DOCUMENT RESUME

ED 057 208 VT 014 229

AUTHOR	Browning, Philip L., Ed.
TITLE	Evaluation of Short-Term Training in Rehabilitation. Oregon Studies in the Rehabilitation of the Retarded.
INSTITUTION	Oregon Univ., Eugene. Rehabilitation Research and Training Center in Mental Retardation.
SPONS AGENCY	Social and Rehabilitation Service (DHEW), Washington, D.C.
REPORT NO	Monog-3
PUB DATE	Aug 70
NOTE	107p.
AVAILABLE FROM	Rehabilitation Research and Training Center in Mental Retardation, University of Oregon, 3rd Floor, Clinical Services Bldg., Eugene, Oregon 97403 (No charge)
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Annotated Bibliographies; Mentally Handicapped; Models; *Program Effectiveness; *Program Evaluation; *Rehabilitation Programs; *Short Courses
IDENTIFIERS	Short Term Training

ABSTRACT

Improving the effectiveness of short-term training for personnel serving the handicapped requires adequate evaluation of rehabilitation programs. This monograph presents a sample of evaluative theoretical positions, methodological approaches, and innovative strategies applicable to the evaluation function. Papers included are: (1) "Evaluation Theory Development" by Marvin C. Alkin, (2) "The Evaluation of Occupational Education Programs" by Jerome Moss, Jr., (3) "Evaluation of Training" by Donald L. Kirkpatrick, (4) "On-Going Program Evaluation" by Bryan Smith, (5) "Behavioral Criteria for Short-Term Training" by Leo A. Hamerlynck, and (6) "Professional Opinions Regarding Curriculum Content in Short-Term Training Programs in Mental Retardation: An Evaluation Survey" by Patrick J. Flanigan. An annotated bibliography provides 95 references on the evaluation of training. (BH)

EDO 57208

EVALUATION OF SHORT-TERM TRAINING IN REHABILITATION

ED0 57208

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Oregon Studies in the Habilitation of the Retarded

EVALUATION OF SHORT-TERM TRAINING IN REHABILITATION

Edited by
Philip L. Browning

Rehabilitation Research and Training Center in Mental Retardation

Monograph No. 3

College of Education
Department of Special Education
University of Oregon
Eugene, Oregon 97403

August, 1970

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Rehabilitation Research + Training Center in Mental Retardation University of Oregon TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

Copyright 1970 by the
Rehabilitation Research and Training Center in Mental Retardation
University of Oregon

The Social and Rehabilitation Service reserves a royalty-free, non-exclusive, and irrevocable license to reproduce, publish, or otherwise use, and to authorize others to use, all copyrightable or copyrighted material resulting from this grant supported research.

This publication was supported, in part,
by Rehabilitation Research and Training Center Grant
16-P-56817/0-05 from Social and Rehabilitation Service, Department
of Health, Education and Welfare, Washington, D. C. 20201.

Preface

This is the third of a series of monographs to be published by the Rehabilitation Research and Training Center in Mental Retardation at the University of Oregon. One of several such Centers which are supported, in part, by grants from the Social and Rehabilitation Service of the Department of Health, Education and Welfare, the Oregon Center conducts research related to the rehabilitation of the retarded and provides training for professional personnel engaged in the rehabilitation process.

This monograph grew out of a training seminar held in February, 1969, at the University of Oregon entitled, "Strategies for the Evaluation of Short-Term Training." The program was designed for rehabilitation personnel engaged in administering short-term training programs. The major objective of the conference was to introduce the trainees to various strategies of program evaluation. The editor wishes to express his appreciation to Dr. Samuel Chyatte, Dr. Windle Dickerson, Dr. Robert Leslie, Dr. Patrick Flanigan, Dr. Bryan Smith, and Mr. Robert Miller, the conference speakers, for their assistance in making the program a successful one.

It would be impossible to express appreciation through individual acknowledgment of all those involved in one way or another in the preparation of this monograph. However, a special thanks goes to Miss Cheryl Mulder who supervised the production of the manuscript. Finally, the editor wishes to express his appreciation to the authors and publishers who gave permission for the reprinting of the materials.

P. L. B.

Foreword

The appearance of a monograph on this subject is a welcomed resource in education for rehabilitation. It focuses on a major concern that has long plagued both trainers and administrators.

Past efforts to investigate the effectiveness of short-term training have been more creditable than credible, for myriad research problems confront the investigator—and almost as many critics. A monograph on the evaluation of training in rehabilitation, therefore, would have been welcomed at any time. Its appearance now is particularly timely, as new approaches to the manpower gap in rehabilitation services turn increasingly to inservice training, work-study programs, and various forms of short-term training.

Difficult as it is, evaluation is necessary for the conduct of any effective training. The evaluation of training that relates directly to a job being regularly performed is particularly difficult. For instance, even when an observable improvement in job performance could be noted, it would remain to be shown that it was a particular period of training that had brought about the change.

Administrators with their necessary concern for the cost-benefit aspects of their programs must inevitably seek evidence that a training program was worth either the cash outlay or loss of time of their staff members. They often turn to the trainees themselves with requests for evaluation: What did you get out of it? Should I send others, too? Or they turn to the trainers: Could you do the same thing in less time? Could you set up a course closer by so we wouldn't have to travel so far? Or could you do something different—or better?

Trainers, on the other hand, are justifiably concerned with the effect of a particular course or a particular approach on the trainees. They ask what did you get out of the course? What changes would you make? When asked for their own evaluation, they often respond with the answers they have received in such polling.

We who are in the training business know all too well that neither these answers nor these questions are adequate for evaluating a training program or even a single course. We know they provide valuable trainee reactions which can help trainers improve the teaching or increase the learning potential. We are well aware that even the most conscientious trainer rarely has time or inclination to go further than this.

One would hope that the material on evaluation described here would en-

courage trainers to attempt more precise studies of their training program; for instance, that they would develop ways of describing the desired training outcomes in less global terms than one usually sees in applications for short-term training funds. What a relief it would be to the grantor of funds for training to have goals set forth in measurable or observable terms, reasonable ones that could be achieved in a two- or three-day course!

An equally difficult research problem we would hope to see encouraged is that of individual response to training. Here one is keenly aware of the administrator's stake in his staff development. We would not look forward to a time in the near future, if ever, when an administrator could have a computer tell him which counselor to send to which course at what time. Still, strategies of evaluation which take seriously individual responsiveness could be of great help for long-range planning for short-term training.

Finally, then, in looking forward to the appearance of this monograph, we would hope that it signals another step toward that long-range planning without which any evaluation of short-term training remains inconclusive. In our mind, it is not the evaluation of a single course or even a specific training and learning that will eventually provide rehabilitation services of the quality we all seek.

Margaret F. Clark
Consultant, Short-Term
Training
Division of Training
RSA, SRS, HEW

Contents

Preface iii

Foreword v

Introduction 1

Evaluation Theory Development 9
Marvin C. Alkin

The Evaluation of Occupational Education Programs 17
Jerome Moss, Jr.

Evaluation of Training 35
Donald L. Kirkpatrick

On-Going Program Evaluation 57
Bryan Smith

Behavioral Criteria for Short-Term Training 65
Leo A. Hamerlynck

Professional Opinions Regarding Curriculum Content in Short-Term Training Programs in Mental Retardation: An Evaluation Survey 69
Patrick J. Flanigan

Evaluation of Training: Annotated Bibliography 89

Introduction

Rehabilitation is a concept which has assumed an increasingly influential role within the helping professions over the past several decades. In alignment with this trend has been the establishment of short-term training programs designed to meet certain manpower needs. The ultimate goal of these programs is to provide more effective rehabilitation services to the handicapped. In order to help maximize the impact of this rapidly expanding area of short-term training activities in rehabilitation, considerably more attention needs to be given to the following general proposition:

> The extent to which effective rehabilitation services are delivered to the handicapped is, in part, a function of the extent to which short-term training is effective.

If the above proposition has any validity, then it becomes the responsibility of professional training personnel to make every effort possible to improve the effectiveness of such programs.

This monograph reflects the position that one such effort definitely should be through the EVALUATION of training. To set the stage for the materials that follow, the terms rehabilitation, short-term training and evaluation are discussed. In addition each of the sources contained within the monograph is briefly introduced.

Rehabilitation

Obermann (1965) defined the term "rehabilitation" as "that activity that is required to assist an [handicapped] individual to move from a status of inadequacy to a status of adequacy" (p. 41). At one time "that activity" or service was provided by a limited number of disciplines which operated in a somewhat segregated manner. The current trend, however, is to integrate the services of the medical, psychological, sociological, educational, and vocational professions to provide maximum rehabilitation (Lofquist, N. D.). Although rehabilitation services were previously limited primarily to handicapped individuals with physically disabling conditions, they have now been expanded to include mental, social, economic, and educational disabilities. The goals of rehabilitation have also undergone dramatic change. When the field was mainly vocationally oriented, the disabled person's inadequacy-adequacy status simply

referred to unemployed-employed. The current view of the status of adequacy, however, is the "total" adaptation of the individual (Strauss, 1965).

In order to truly capture one's imagination regarding the growth and expansion of rehabilitation, it would be necessary to describe significant developments and trends over the past several decades. However, selected figures regarding fiscal appropriations for the federal-state vocational rehabilitation program and the number of disabled clients successfully rehabilitated through that program affords some appreciation of the magnitude of change in the field. Annual appropriations of $796,000 and $400 million were provided for rehabilitation efforts for the years of 1920 and 1968, respectively. As for the number of individuals served, only 532 clients were rehabilitated the first year of the federal-state program whereas 208,000 people were successfully rehabilitated in 1968.

In spite of the tremendous gains made in the field of rehabilitation there are certain problems which continue to prevent maximum rehabilitation services. One of the most urgent and critical of these problems is that of manpower. As recently as 1968 a National Advisory Committee issued the following statement:

> We believe that appropriations for support of training should be substantially increased if they are to be even nearly commensurate with the magnitude of manpower needs and the resources of educational institutions and cooperating agencies for the preparation of personnel in the professional fields most closely related to rehabilitation of disabled people. (Report of the National Citizens' Advisory Committee on Vocational Rehabilitation, 1968, p. 56.)

The first major attempt to confront the manpower problem was in 1954 through Public Law 83-565 which provided for grants to support the training of more professional personnel for rehabilitation services. Of five types of grants provided, one was for short-term training.

Short-Term Training

This type of training can be viewed in several ways; for our purpose, however, the following definition is offered:

> Training that is of shorter duration than a quarter or a semester of a regular academic year is generally regarded as short-term training. It is intended to precede or supplement basic or advanced professional education on an academic or calendar year basis. Short-term training courses generally range in length from 2 or 3 days to 6 weeks, according to their specific purpose. They include institutes, workshops, seminars, and other training courses (Support of Short-Term Training, 1966, p. 48).

When the training program commenced in 1954, $900,000 was appropriated for training grants. In 1967 a total of $29,700,000 was appropriated (Report of the National Citizens' Advisory Committee on Vocational Rehabilitation, 1968) with a substantial portion of this designated for short-term training grants.

A major objective of the short-term training program has been to provide rehabilitation personnel already in the field with a better understanding of the methods and philosophy of rehabilitation, as well as to afford them an opportunity to raise their level of knowledge and skill. In essence, short-term train-

ing is one means of attempting to facilitate the effectiveness of rehabilitation services provided by personnel serving the handicapped.

The scope of the program is partly reflected by the disciplines it serves. For example, short-term training grants are now made available to the fields of physical medicine and rehabilitation, physical therapy, occupational therapy, speech pathology and audiology, rehabilitation nursing, rehabilitation social work, prosthetics and orthotics, rehabilitation psychology, rehabilitation counseling, recreation for the ill and handicapped, as well as other specialized fields.

In 1959 short-term courses and institutes reached only 2,500 professional personnel already in the field; in 1968 the Rehabilitation Services Administration funded this type of training so that more than 9,000 individuals were served (Hunt, 1969). In addition, the Regional Rehabilitation Research and Training Centers in 1967 conducted 364 short-term training courses which served 18,415 trainees (Report of the National Citizens' Advisory Committee on Vocational Rehabilitation, 1968). These figures do not include the short-term training courses offered through the states' rehabilitation in-service training programs.

Of course, the extent to which the short-term training programs adequately meet certain manpower needs depends on the effectiveness of the individual programs. Thus, it becomes the responsibility of each training director or persons in charge of short-term courses to make every effort to improve their program so that the trainees will be able to provide more effective rehabilitation services to the handicapped. As already indicated, one such effort definitely should be through the evaluation of training.

Evaluation

Although attempts to treat this term have been many and varied, a major failing in evaluation has been said to be a lack of adequate definition (Alkin, 1969). The following definition is presented primarily because it recognizes evaluation as a process that involves the identification of many factors that contribute to educational or training outputs.[1]

> Evaluation is the process of first identifying and then quantifying or measuring the relationships between student [trainee] inputs and educational [training] outputs and determining the combination of mediating factors which maximizes the educational [training] outputs, given a constant financial input and controlling for the effects of external systems (Alkin, 1968, p. 1).

Trainee inputs refer to the nature and characteristics of those entering a training program, e.g., type and level of professional training, motivation. Alkin speaks of educational or training outputs as referring to (a) both cognitive and non-cognitive changes that the person experiences following an instructional program, and (b) program impact upon systems external to it, e.g., agency, community.

Mediating factors are descriptive characteristics of the program itself, e.g., personnel, curriculum, small groups, audio-visual aids, practicum experience. Financial inputs simply refer to the monetary resources made available for carrying on the program. Finally, external systems mean the framework of

[1] The reader is referred to Moss' and Alkin's articles in this booklet for two additional definitions of evaluation.

social, political, legal, economic, and other systems outside the training program which encompass the program, have impact upon it, and are in turn modified by the outputs of the program. The reader is referred to Alkin's (1968) paper for further discussion and a figurative presentation of this definition. Since there is a diversity of opinion regarding what constitutes evaluation, this definition represents only one of many and varied viewpoints. General agreement probably exists, however, as to the important purpose it serves in a training program.

Tracey (1968) stated that evaluation served to determine:

1. . . . where the activity is at any given moment and providing a baseline for measuring progress.
2. . . . the value of training and development program activities to the enterprise and of appraising the efficiency and effectiveness of the functions performance of the task set forth.
3. . . . whether the time, energy, and money expended in planning and operating programs of training and development are producing results sufficient to justify the investment (pp. 12-13).

Tracey also listed three functional ways in which an adequate evaluation program is critically important to training:

> *First*, the steady growth of training and development activities in most enterprises, which in total involve millions of people and many more millions of dollars, makes it essential that those responsible for the management of these activities be able to defend their programs by knowing the accomplishments and contributions of the activities to the enterprise goals. The continued support of top management hangs in the balance. *Second*, evaluation provides trainers with a means of determining the efficiency, effectiveness, and utility of both management and operation. Only by appraisal is it possible to ensure that programs are suited to the groups for which they are designed and that they result in the behavioral changes required for improved products or services. *Third*, evaluation provides a starting point for the design of an improvement program (p. 13).

Finally, he set forth several fundamental assumptions which underlie the need for evaluating training:

1. Any training and development program must be validated; that is, the efficiency and effectiveness of programs must be objectively determined. They must be subjected to critical evaluation and must demonstrate their value to the organization if they are to be retained.
2. Any training or development program can be improved—no program is perfect. Although the effectiveness of the program may have been demonstrated, further refinements are possible.
3. Improvement of any training or development program can be affected by:
   a. objective and coordinated evaluation of every aspect of the operation.
   b. the application of imagination and creative thinking by all personnel.
   c. deliberate collection of the observations, ideas, and thinking of all personnel.

d. critical analyses and synthesis of findings, ideas, and alternatives.
e. systematic, time-phased development and tryout (Tracey, 1968, pp. 13-14).

If short-term training is going to have a significant influence, then greater attention certainly must be given to the evaluation of such training activities.

OVERVIEW OF THE MONOGRAPH

This monograph is not intended to serve as a comprehensive overview of evaluation. Consequently, the reader is presented with only a sample of evaluative theoretical positions, methodological approaches, and innovative strategies. Nevertheless, it is hoped that what is contained within these covers will serve as an "eye opener" and an incentive to more and better evaluation of short-term training in rehabilitation.

Marvin C. Alkin, *Evaluation Theory Development*

Alkin presents only one of a number of evaluation theories or models which have been offered. In addition to being introduced to an evaluation model which is couched in a decision-making framework, the reader is presented a definition of evaluation and five evaluation need areas.

Evaluation is a complex process which can represent the interrelationship of an infinite number of parameters or variables. Thus, evaluation theories, systems, or models are needed to enable us to reduce the complexities of evaluation to more manageable proportions.

Jerome Moss, Jr., *The Evaluation of Occupational Education Programs*

Moss presents an evaluation model which is counter to the often raised but somewhat misleading question, "Is short-term training effective?" This type of question presents a number of difficulties. In the first place, it fails to consider the input of differential trainee variables. Secondly, it often views training (the independent variable) as a uni-dimensional entity. A more meaningful question that training personnel should be asking, and one upon which Moss' model reflects, is "What kind of trainees (input), given what kinds of training experiences (process), demonstrate what kinds of change (output)?" In addition to the differential evaluative model, Moss discusses a number of important considerations in evaluation, e.g., definition, outcome criteria, types of research approaches.

Donald L. Kirkpatrick, *Evaluation of Training*

The author discusses an approach to evaluation which involves what he proposes to be the four major steps in the evaluative process, i.e., reaction, behavior, learning, and results. In addition, the reader is provided with guidelines, procedures and examples of the four steps. Even though the examples are taken from the fields of business and industry, it is important to stress that the guidelines, procedures and techniques described are certainly applicable to the evaluation of short-term training in rehabilitation.

Bryan Smith, *On Going Program Evaluation*

Evaluation is often viewed as being limited to a pre-post test approach for

the purpose of assessing the effects of training. However, most experts in the field now take the position that evaluation is an on-going process and that the program planning phase is one of the more important evaluative stages to be considered. Beckard (1962) stated, "all planners can improve the effectiveness of their meetings by building into their planning systematic ways of appraising the situation as they go along (p. 52)." Smith presents the reader with a systematic tool for program planning and on-going evaluation—Program Evaluation and Review Technique (PERT).

Leo A. Hamerlynck, *Behavioral Criteria for Short-Term Training*

The type of outcome criteria we choose to select for evaluating the effectiveness of short-term training in rehabilitation is of critical importance. This author presents a categorization model which includes four types of criteria and discusses the implications of each. The direct-primary type (actual client behavior) of outcome criteria has been ofen ignored yet is most germane since the ultimate purpose of rehabilitation short-term training is to better help the handicapped.

Patrick J. Flanigan, *Professional Opinions Regarding Curriculum Content in Short-Term Training Programs in Mental Retardation: An Evaluation Survey*

The author presents a descriptive summary of a curriculum evaluation for rehabilitation short-term training programs in mental retardation. The survey was based, in part, on the assumption that enrichment and modification of short-term training programs are most realistically brought about by consultation with professional people in the field. Over a four year span, 3,299 practitioners representing 12 professional orientations were screened regarding their suggestions for specific curricula which they thought was applicable for short-term training programs in mental retardation.

Evaluation of Training: Annotated Bibliography

Ninety-five annotated references are provided on the evaluation of training, most of which have been derived from the fields of business and education. The reason for this is that the rehabilitation literature is practically nil on the subject with the exception of the brief attention given to the area by the Institute on Rehabilitation Services (IRS).

Personnel in the fields of business and industry have been largely responsible for the advancements made in training technology, and have also been leaders in the area of evaluation of training. As well, evaluation specialists are now emerging in the field of education which is indicated partially by the UCLA Center for the Study of Evaluation and the Ohio State University Evaluation Center. The time has come when rehabilitation training personnel must begin borrowing training assessment strategies from these and other professional fields.

References

Alkin, M. C. Towards an evaluation model: A systems approach. University of California, Los Angeles: Center for the Study of Evaluation of Instructional Programs. Working Paper No. 4, February, 1968.

Alkin, M. C. Evaluation theory development. *Evaluation Comment*. UCLA Center for the Study of Evaluation, 2, No. 1, October, 1969.

Beckhard, R. Evaluating the effectiveness of meetings. In R. Beckhard (Ed.) *Conference planning: Selected readings series–six*. Washington, D.C.: National Training Laboratories, National Education Association, 1962.

Hunt, J. A decade of progress. *Journal of Rehabilitation*, January-February, 1969, 9-12.

Lofquist, L. H. Introduction. In L. H. Lofquist (Ed.) *Psychological research and rehabilitation*. Washington, D.C.: American Psychological Association, N.D.

Obermann, C. E. *A history of vocational rehabilitation in America*. Minneapolis: T. S. Denison and Company, Inc., 1965.

Report of the National Citizens' Advisory Committee on Vocational Rehabilitation, *National citizens' advisory committee on vocational rehabilitation*. Washington, D.C.: U. S. Government Printing Office, June 26, 1968.

Strauss, R. Social change and the rehabilitation concept. In M. B. Sussman (Ed.) *Sociology and rehabilitation*. American Sociology Association, 1965.

Support of short-term training. *Vocational rehabilitation administration training grant program*. Washington, D.C.: U. S. Department of Health, Education, and Welfare, 1966.

Tracey, W. R. *Evaluating training and development systems*. American Management Association, Inc., 1968.

Evaluation Theory Development[1]

Marvin C. Alkin, University of California, Los Angeles

The development of a theory of evaluation has been established as the major goal of the Center for the Study of Evaluation. A theory of evaluation should: (1) offer a conceptual scheme by which evaluation areas or problems are classified, (2) define the strategies including kinds of data, and means of analysis and reporting appropriate to each of the areas of the conceptual scheme, and (3) provide systems of generalizations about the use of various evaluation procedures and techniques and their appropriateness to evaluation areas or problems.

At their best the propositions presented in a theory of evaluation should enable one to predict, fully, the appropriateness of utilizing various evaluation strategies within a system. Development of an evaluation theory is thus an "end" rather than a means, guiding the research activities of the Center for the Study of Evaluation. Development of a theory is a difficult enterprise. The process of working toward the achievement of this end requires a conceptual framework to guide and coordinate our efforts. What is presented in this paper represents months of conceptual efforts and may be thought of as a first approximation of an attempt to develop an evaluation theory. In other words, we have constructed a rationale for conducting evaluations in a certain way that is based on a specified set of assumptions which in turn underlie a precise definition of what an evaluation is supposed to do or be.

To start with, in the development of a theory, it is necessary to reach agreement on a definition of evaluation. Most would agree that a major failing of evaluation today stems from the lack of an adequate definition. Past definitions of evaluation have either equated it with: (1) measurement and testing, (2) statements of congruence between performance and objectives, or (3) professional judgments. None of these definitions by itself is sufficient to provide all the necessary information or to include the multiplicity of activities now regarded as evaluation.

In the past year, there has been increasing evidence of a developing consensus on a broader, more comprehensive definition of evaluation. This ex-

[1] Published in *Evaluation Comment,* October 1969, *2,* (1), 2-7, and reprinted with the permission of the author.

panded view has necessarily taken into consideration that the judgments from evaluators are intended to be of use to decision-makers in selecting among various courses of action. This view of evaluation also acknowledges the uniqueness of specific situations or programs and the necessity of recognizing this uniqueness in the evaluation as well as in the manner in which the evaluation information is ultimately reported.

1. Evaluation is a process of gathering information. Most past definitions of evaluation are inadequate since they do not cover the full range of activities requiring information.
2. The information collection in an evaluation will be used mainly to make decisions about alternative courses of action, rather than being employed in some other fashion. Thus, the manner in which the information is collected, as well as the analysis procedures, must be appropriate to the needs of the decision-maker or of potential decision-involved publics. This requirement might necessitate quite different analyses than those which might be employed if the purpose were understanding the education process per se.
3. Evaluation information should be presented to the decision-maker in a form that he can use effectively and which is designed to help rather than confuse or mislead him.
4. Different kinds of decisions may require different kinds of evaluation procedures.

While there are any number of variations of a specific wording that might serve equally well for a definition of evaluation, we have devised one which fits our conceptions of evaluation and meets our biases. We would maintain that evaluation must take into consideration the ultimate decision-making functions to be served, as well as the nature of the specific problem or situation under analysis. We prefer the following definition:

> *Evaluation is the process of ascertaining the decision areas of concern, selecting appropriate information, and collecting and analyzing information in order to report summary data useful to decision-makers in selecting among alternatives.*

The first part of the definition of evaluation presented here deals with *ascertaining the decision areas of concern.* The decision-maker, and not the evaluator, determines the nature of the domain to be examined. The evaluator can and should, however, point out inconsistencies, potential difficulties, or additional data that might modify the decision-maker's views on the relevance of certain concerns.

For example, if the evaluator is called upon by a specific decision-maker to provide an evaluation, first, he will want to know what should be evaluated. Decision areas of concern may be stated relative to explicit statements of goals or objectives of the system or relative to various implicit goals. In his interactions with the decision-maker, the evaluator may wish to point out the necessity for broadening the area of concern because of interrelated aspects of the school [training] program, or to consider, as well, various areas of potential unanticipated outcomes.

On the other hand, if the evaluator is conducting an evaluative study of an educational institution without having been commissioned by a specific decision-maker, he has available greater flexibility. A professor, for example,

might conduct an evaluative study of his university. There is a preconception on the part of the evaluator as to which decision-maker or potential decision groups he is directing his work towards. Thus, the decision area of concern in such an endeavor is framed by the unique nature of the potential decision-maker or decision groups along with either actual data or judgments on the part of the evaluator as to the concerns of this group (individual).

We consider this "preconception of decision-maker" notion a fundamental and useful distinction between evaluation and some kinds of research. If one realizes that the purpose of what he does is to provide the best possible basis for informed judgments or decisions, his thinking about his task will surely be influenced; and this will be a different influence than that which operates on the researcher whose purpose is to discover or explain some phenomenon.

Another part of the definition and, therefore, another task of evaluation deals with *selecting appropriate information* in light of the decisions areas to be considered. If the decision area relates to the assessment of the needs of a total system, the information requirements will be quite different than when the decision area is related to the relative success of two specific alternative programs conducted under experimental conditions. The task of the evaluator in specifying information requirements includes the development of the evaluation design of the project, and the selection and/or development of instruments designed to provide the information appropriate to the decision areas.

Collecting and analyzing the information are tasks of prime concern to the evaluator. He will encounter different problems associated with these tasks, depending upon the unit being evaluated, the nature of the decision-maker, and other considerations.

One of the most vital parts of the evaluation process is *reporting summary data* to the decision-maker. Most evaluators often overlook this function as being merely a pro forma exercise. The evaluator's role requires that he make judgments about the relative worth of various courses of action. These judgments may be in the form of statements or recommendations to the decision-maker(s), or may be general descriptive material. But in all instances the evaluator should attempt to be explicit in the specification of the value system that led to the judgments made. Indeed, if the purpose of evaluation is to provide information that will enable decision-makers to reach decisions about alternatives, then the nature and form of the reporting should be appropriate to the problem and the audience.

The summary data is provided to be of use to the *decision-maker*. It has already been alluded to in this paper that we are using the term "decision-maker" to apply both to an explicit contractor of evaluation services as well as a potential but only implicit decision-maker or group. Moreover, we are using the term "decision-maker" to apply both to an individual with organizational "line" authority (e.g., a school principal [a training director]) as well as to other publics that participate in the decision process or in the development of educational policy decisions. Throughout this paper, whenever we refer to "decision-maker" it is in the generic sense discussed above.

Information is provided to "decision-makers," in order to enable sounder decisions in *selecting among alternatives*. By definition, a "decision" involves making a choice among alternatives. However, the form of alternatives has a wide range. Alternatives may range from a "go/no-go" category regarding a given textbook for a particular classroom to a complex aggregation of a number of budget categories related to an optimum expenditure level. In general, the number of categories of alternatives increases as the size of the

program or system increases; e.g., pupil achievement, teacher morale, teacher practices, etc.

The summary evaluation data should ordinarily be presented in the form of statements and/or recommendations about alternatives. An exception would be when such information is designed to describe the status (past, present, or future) of the system. For example, "the students at your school are weak in mathematics." In this instance, there are no alternatives and the decisions are implied (e.g., something should be done to correct this situation).

EVALUATION NEED AREAS

The foregoing definition and assumptions are closely tied to the decision-making process, which in turn leads to a consideration of what kinds of educational decisions require evaluative information. Inquiry along these lines has led to the development of a decision-oriented classification of the various types of evaluation. Five areas of evaluation may be identified.

These five areas represent attempts to provide evaluative information to satisfy unique decision categories. In other words, there are evaluations necessary in providing information for decisions about the state of the system. (We call such evaluations *systems assessment*.) There are evaluations necessary in providing information to assist in the selection of particular programs likely to be effective in meeting specific educational needs. (We call this kind of evaluation, which takes place prior to the implementation of the program, *program planning*.) There are evaluations necessary in providing information relative to the extent to which a program has been introduced in the manner in which it was intended and to the group for which it was intended (*program implementation*). There are evaluations necessary in providing information during the course of a program about the manner in which the program is functioning, enroute objectives are being achieved, and what unanticipated outcomes are being produced. Such information can be of value in modifying the program (*program improvement*). Evaluations are necessary in providing information that might be used by decision-makers in making judgments about the worth of the program and its potential generalizability to other related situations (*program certification*).

The evaluation areas outlined above seem to represent a growing consensus among a number of people engaged in the study of evaluation. The first two and the last need areas discussed are somewhat similar, respectively, to "context," "input," and "product" presented by Stufflebeam (1968). What he refers to as "process" we have chosen to think of as two separate stages, program implementation and improvement. Major differences of emphasis are found between the descriptions of our need areas and his stages. However, Stufflebeam's work contributed substantially to our thinking. The discrepancy model presented by Provus (1969) outlines five stages: "definition," "installation," "process," "product," and "cost benefit analysis." The first four of these are somewhat similar to our stages two through five. Provus does not include systems assessment as a part of his model. With respect to cost benefit analysis, we would maintain that cost benefit considerations are a part of each need area of the evaluation and cannot be thought of as simply an additional task to be attacked when all of the other evaluations have been completed. The general notion that cost benefit considerations are a part of each stage of the evaluation process also would seem to be subscribed to by John Hemphill in his recent NSSE Yearbook chapter (1969). The evaluation types outlined by

Rodney Skager (1969) are also quite similar to the framework presented here. A better understanding of the need areas will be gained by a more complete description of each.

Systems Assessment

Systems Assessment is a means of determining the range and specificity of educational objectives appropriate for a particular situation. The needs may be represented as a gap between the goal and the present state of affairs. The evaluative problem then becomes one of assessing the needs of students, of the community, and of society in relation to the existing situation. Assessment, therefore, is a statement of the status of the system as it presently exists in comparison to desired outputs or stated needs of the system.

A systems assessment might be related to evaluation of a specific instructional program and thus the charge would be to determine the present status relative only to a specific objective and related objectives. We would refer to this as a "sub-system assessment."

Systems assessment does not refer to specification of process characteristics appropriate for a district, school, or classroom. A statement such as "this district needs a lower pupil-teacher ratio" or "a need of this district is to install team teaching" is not a systems assessment. The systems assessment must be related to the ultimate behavior of clients of one type or another (pupils, parents, community, etc.—all clients of the school). To put it simply, systems assessment must result in a statement of objectives in terms of outputs of the school [training program].

The process in the systems assessment area of ascertaining the decision area, specifying and collecting information and reporting summary data, requires methodology and techniques different from that which might be employed, for example, in a typical experimental design. The data are concerned with the status of the system. The summary data might be comparative, historical, or other descriptive information.

Program Planning

Program planning, the second need area, is concerned with providing information which will enable the decision-maker to make planning decisions—to select among alternative processes in order to make a judgment as to which of them should be introduced into the system to fill most efficiently the critical needs previously determined. In an instance where we are proceeding through severe need areas in sequential fashion, the following might occur. After the decision-maker receives the systems assessment evaluation, he might make a decision as to the appropriate means of fulfilling that need. Alternatively, he might designate several possibilities and ask the evaluator to provide information on the possible impact of each. Hence, in program planning, the evaluator provides the data for an evaluation of a program prior to its inception. The task of the evaluator is to anticipate the attainment of goals and to assess the potential relative effectiveness of different courses of action.

It is quite obvious that the collection and analysis of data of the type required for this evaluation need area will be quite different from collection and analysis problems for other areas. The techniques may require both internal and external evaluation procedures. (See Lumsdaine, 1965.)

By way of internal evaluation, programs may be examined to determine the extent to which their reproducible segments purport to achieve the objectives of the program being evaluated. Technical features of style or construction,

practicality and cost are other means of providing internal evaluation. To date, the evaluations of products by EPIE have been primarily based upon internal evaluations.

External evaluations of programs yet to be implemented might take the form of examining research data on the results of implementation in similar or near-similar situations. Or external evaluations might attempt to utilize some of the various educational planning techniques to obtain data. Computer simulations might be developed; Delphi analysis might provide insights into the potential outcomes of a program; gaming and various other systems analytic approaches might also provide external evaluation data.

Program Implementation

After the decision-maker has selected the program to be implemented, an evaluation of program implementation determines the extent to which the implemented program meets the description formulated in the program planning decision. In the case of an existing program where no known changes have been implemented, the evaluation task at this stage is to determine the degree to which planning descriptions of the program coincide with the implemented program and the extent to which assumed descriptions of inputs to the system (students) correspond with observed inputs.

There have been numerous examples in the educational literature of conflicting results relative to the impact of a specific instructional treatment. We would maintain that in large part this is attributable to the lack of specificity of the precise nature of the instructional treatment that was employed. Team teaching is *not* always team teaching. More precisely, team teaching in Santa Rosa might be quite different from team teaching in Boston or from team teaching in Palo Alto, California. The precise definition or the parameters defined as team teaching in a given situation would help to insure an understanding of what is being evaluated and whether what is being evaluated is what the investigator thought the program was.

Program Improvement

The evaluator can play an important role in program improvement, the fourth need area, by providing as much information as possible about the relative success of the parts of the program. In order to perform program improvement evaluation, it is necessary to recognize the basically interventionist role that the evaluator has been asked to take.

The key point in the understanding of the role of the evaluator in performing evaluations in this need area is that he is first and foremost an interventionist attempting to provide data which will lead to the immediate modification and, hopefully, improvement of the program. As the evaluator identifies problems and collects and analyzes related information, data are presented immediately to the decision-maker so that changes may be executed within the system to improve the operation of the program. Information might include data on the extent to which the program appears to be achieving the prescribed objectives, as measured by regular tests; information also might be presented which relates to the impact of the program on other processes or programs.

This need area has often been overlooked or ignored by the traditional evaluator who has attempted to impose the antiseptic sterility of the laboratory on the real world. Such an approach may make for a fine experiment, but it does little to improve a program which is often not in its final form.

Program Certification

In the fifth evaluation need area, program certification, the role of the evaluator is to provide the decision-maker with information that will enable him to make decisions about the program as a whole and its potential generalizability to other situations. The evaluator might attempt to provide information which will enable the decision-maker to determine whether the program should be eliminated, modified, retained, or introduced more widely.

The kind of information collected for program certification decisions is in large part dependent upon who is the intended decision-maker. It is obvious that different information will be required if the potential decision-maker is the teacher, the principal, or a funding agency. Evaluations in this area will be concerned with examining the extent to which the objectives have been achieved, as well as with the impact on the outcomes of other programs.

In program certification evaluations, there is a requirement for valid and reliable data which would generally require that the evaluator attempt to apply as rigid a set of controls as possible. The evaluator might use pre- and post-test designs and employ sophisticated methods for analyzing the data. Intervention should be avoided in evaluations in this need area. Here the traditional evaluator is "at home."

In considering the situations in which evaluation might take place in various need areas, we have found it helpful to differentiate between the evaluation of educational systems and the evaluation of instructional programs. In terms of the conceptual framework that has been presented, one can view the evaluation of educational systems as involving the first two need areas and the evaluation of instructional programs as largely involving the last three.

In evaluating any educational system it is necessary to determine the educational needs in terms of the most appropriate objectives for the given system and to devise a procedure for providing regular information on the progress of the system relative to these dimensions. This procedure is the evaluative device for decision-making about the assessment of system needs (Systems Assessment). When decisions have been made about the objectives of the system which are inadequately met, the decision-maker might then be concerned with the selection of programs to meet these objectives. Evaluation information might be sought relative to the possible impact of various courses of action or programs (Program Planning).

Thus, if one followed through on the full cycle of evaluation in an educational system, including the allowance of feedback and recycling, the process might be depicted as in Figure 1.

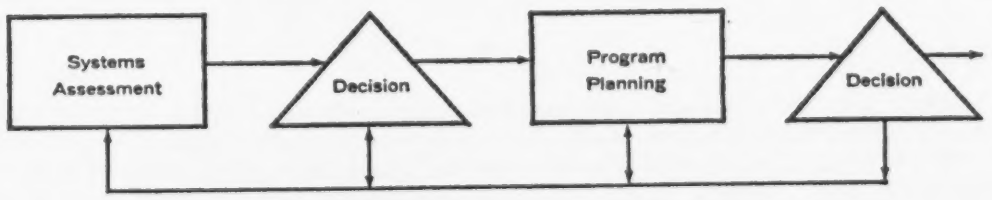

Fig. 1. Evaluating Educational Systems

The evaluation of an instructional program assumes the prior assessment of the program or of a larger system, a decision about objectives to be attended to, and the selection of programs considered to be appropriate for meeting these objectives. That is, the evaluation of an instructional program ordinarily begins after the decisions related to need areas 1 and 2 of the evaluation have

been made. In evaluating an instructional program, the objectives to be achieved and the program which it is assumed will be most successful in achieving these objectives are generally considered as "given." Thus, the evaluation of an instructional program focuses primarily on the last three need areas of evaluation.

Where the evaluation task commences with the evaluation of the instructional program, we envisage the necessity for a sub-system assessment dealing with the area of concern of the selected instructional program. Thus, it is seen that the evaluation need areas are not necessarily sequential with the steps easily defined. In some instances, moreover, the data collection, analysis, and reporting appropriate to a decision might be so easy to obtain or so inextricably tied to the making of the decision that the decision-maker and his staff would perform the evaluation themselves. In some instances, the project begins for the evaluator after a number of decisions have already been made. Thus, the evaluator might have to attend to only selected evaluation need areas.

For the sake of convenience, in Figure 2 we have depicted a way in which the evaluation need areas might be interrelated in the evaluation of instructional programs.

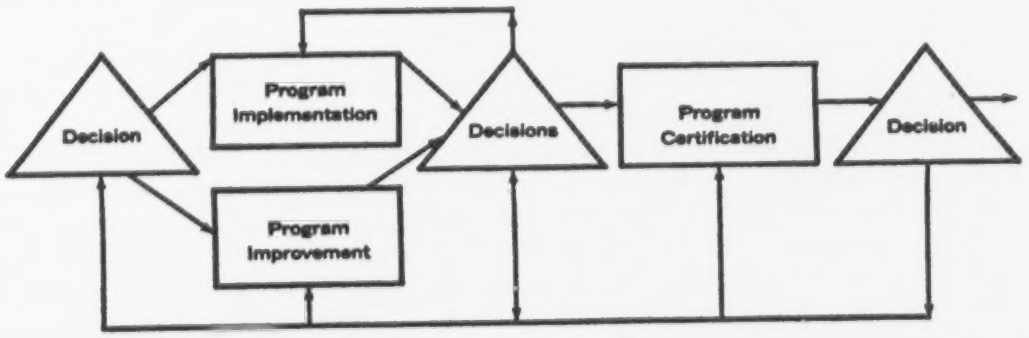

Fig. 2. Evaluating Instructional Programs

A final explanatory note is in order concerning the role of the evaluator in this evaluation model. It might be possible to draw the conclusion that the evaluator does all things—that he is curriculum designer, administrator, program implementor, test officer, budget manager, etc. This is a misconception. We have partially dispelled this notion by commenting earlier that what has been described in this section is the full range of the evaluation cycle. We are describing functions to be performed rather than a role in each evaluation need area for a specific individual.

References

Hemphill, J. K. The relationships between research and evaluation studies. In R. W. Tyler (Ed.) *Educational evaluation: New roles, new means; the sixty-eighth yearbook of the national society for the study of education.* Part II. Chicago: NSSE, 1969, 189-220.

Lumsdaine, A. A. Assessing the effectiveness of instructional programs. In R. Glaser (Ed.) *Teaching machine and programmed learning.* Vol. II: *Data and directions.* Washington, D.C.: National Education Association, 1965, 267-320.

Provus, M. Evaluation of ongoing programs in the public school system. In R. W. Tyler (Ed.) *Education evaluation: New roles, new means; the sixty-eighth yearbook of the national society for the study of education.* Part II. Chicago: NSSE, 1969, 242-283.

Skager, R. W. *Evaluation and the improvement of compensatory educational programs.* CS7 Report No. 54, Los Angeles: Center for the Study of Evaluation, 1969.

Stufflebeam, D. L. *Evaluation as enlightenment for decision-making.* Columbus: Evaluation Center, Ohio State University, 1969.

The Evaluation of Occupational Education Programs[1]

Jerome Moss, Jr., University of Minnesota

The justification for this paper rests upon the validity of three assumptions: program evaluation is essential to systematic improvement in educational efficiency and effectiveness; an intensification of evaluative activity is highly desirable; much of what little has been done to date in the name of program evaluation is of questionable usefulness.

Consistent with the assumptions, the purposes of the paper are to stimulate more productive studies by exploring some of the dimensions of program evaluation and to provide a conceptual framework for evaluative efforts in vocational, technical, and practical arts education.

The paper will treat the following eight dimensions of evaluation, plus a brief section dealing with some implications for action: (a) the importance of program evaluation, (b) some causes of past inactivity in evaluation, (c) a definition of program evaluation, (d) program outcomes (or evaluative criteria), (e) program characteristics, (f) two roles of program evaluation, (g) evaluation as a part of the educational change process, and (h) some research approaches to evaluation.

THE IMPORTANCE OF PROGRAM EVALUATION

A basic premise of the paper is that program evaluation is important because it provides evidence about the relative merits of programs, thus enabling educators to make more rational decisions about the theories and practices of program development and operation. Improved rationality of decision-making is demanded on moral, social, and scientific grounds.

We have a moral obligation to students to provide them with the best programs possible. But first, we must be able to identify the best of those available.

[1] Published in *Technical Report*, September 1968, Research Coordination Unit in Occupational Education, University of Minnesota, and reprinted with the permission of the author.

We must be able to show, for example, whether or not, and in what measure, a newly developed program is superior to the standard program.

We have a social obligation. As the public investment in vocational, technical, and practical arts education rises, it becomes crucial that the money is spent with the greatest efficiency for society's ultimate welfare.

We have a scientific obligation. Evaluation is an indispensable part of the scientific method. Educators hypothesize that programs with certain characteristics will yield certain outcomes. They develop programs with those characteristics and try them out. The measurement of outcomes to confirm or deny the hypotheses is necessary to produce verifiable knowledge. Evaluation is thus essential to the development of a science of instruction, without which we shall continue to operate by hunches, authority, tradition, and personal experience.

One can be against evaluation, therefore, only by showing that it is improper to seek an answer to the question about the relative merits of educational programs. It is entirely fitting and common to criticize the way in which the evaluation is done and the validity of the outcome, but it is hardly ever appropriate to argue against the need for evaluation or the importance of improving evaluative techniques.

SOME CAUSES OF PAST INACTIVITY IN EVALUATION

In light of its importance, how can the fact be explained that program evaluation in vocational, technical, and practical arts education [and rehabilitation short-term training] has been an incidental, casual, and sporadic activity? As a matter of fact, why have relatively few evaluative studies been conducted since the passage of the Vocational Education Act of 1963?

There are philosophical-political reasons. We are faced with critical manpower and social problems. The public policy decision has been to divert most available human and financial resources from both research and evaluation into the most visible approaches that can at least sustain the illusion of progress. With most public agencies under pressure to produce immediate results, it is no wonder that the need for a good show often overwhelms scientific objectivity; it is not surprising that there is little time to revise, throw out, and frankly compare. Careers are often at stake. Further, it is somehow un-American to be indefinite and doubtful, or to adopt a try-and-see attitude about any proposed public program. Legislators are loathe to provide large sums of money to try out several alternative solutions; we pick one "solution" and go. Obviously, evaluation is done hesitantly, with very grave consequences usually associated with unfavorable findings. "When ideas that are promising as objects of research and honest experimentation [are accepted prematurely and] give birth, through artificial dissemination, to a brood of hysterical fads, there is the danger that angry reaction will dump out the egg with the shell" (Oettinger, 1968, p. 76-77).

There are personnel reasons for the relative inactivity. Vocational, technical and practical arts education has been handicapped by a shortage of well-trained researchers, and evaluation has not been looked upon (mistakenly from my perspective) as a specially rewarding, creative form of research.

Finally, there have been (and still are) technical difficulties. The remainder of this paper will touch on some of these problems so they will not be enumerated here. Suffice it to say for the moment that evaluation is a highly complex, technically and conceptually demanding activity. Until relatively recently,

we lacked the statistical and computational tools necessary to do a reasonable job.

From now on, however, the picture must change. The Advisory Council on Vocational Education has demanded greater efforts at evaluation. Social scientists from a wide variety of disciplines are turning their attention to the assessment of various systems of manpower training. Our social obligation for evaluation is being assumed by others, and the results could determine our very existence. We must evaluate our own programs using appropriate criteria and methodology so that decisions concerning our future can be based upon data which properly reflects our educational perspectives.

A DEFINITION OF PROGRAM EVALUATION

While all vocational, technical, and practical arts educators have some concept of program evaluation, the literature indicates that the meanings held are quite varied. Figure 1, entitled "Major Components of the Evaluative System," is intended to introduce the concept as it is used in this paper.

The evaluative system starts with students, each of whom differ with respect to characteristics which affect their ability to learn at the time they enter the program to be evaluated. Students differ, for example, in relevant aptitudes, achievement, motivation, health, etc., which alone and in interactions, create variation in "readiness" for the program.

The program the students enter has characteristics which provide them with educational experiences. It is these characteristics that are to be evaluated. Students are exposed to selected content, which has been organized in specific ways, which is presented in certain manners, and to which the students are encouraged to respond in particular ways, all under the guidance or management of an instructor with certain characteristics. These "transactions" (Stake, 1967) take place under particular physical and psycho-social environmental conditions.

In addition to the influences of the specific program to be evaluated, students are inevitably affected by other experiences and conditions in the environment, which occur outside of the program, but whose effects might be mistaken for outcomes of the program. These experiences can take place at any time after the student enters the program, and before the program outcomes are measured. For example, students might take a variety of other courses which differentially alter their ability to learn the content of the program to be evaluated; increases in dependents or extra-curricular experience on a part-time summer or after-school job could change motivation; economic conditions could alter the availability of particular kinds of jobs after graduation; military service could result in greatly enhanced occupational skills, etc.

The interaction of student characteristics, program characteristics, and other intervening influences produce actual outcomes. These outcomes consist of student or ex-student behaviors, and the effect of those behaviors on the school, the community, the economy, society, etc., and other direct consequences of the program for teachers, administrative patterns, other students, etc.

Finally, the evaluative system contains one or more sets of comparative outcomes. These outcomes are anticipated, expected, hoped-for results of the program, or they may be the actual outcomes of a different program, or the outcomes of the same program at different points in time. In all cases, they

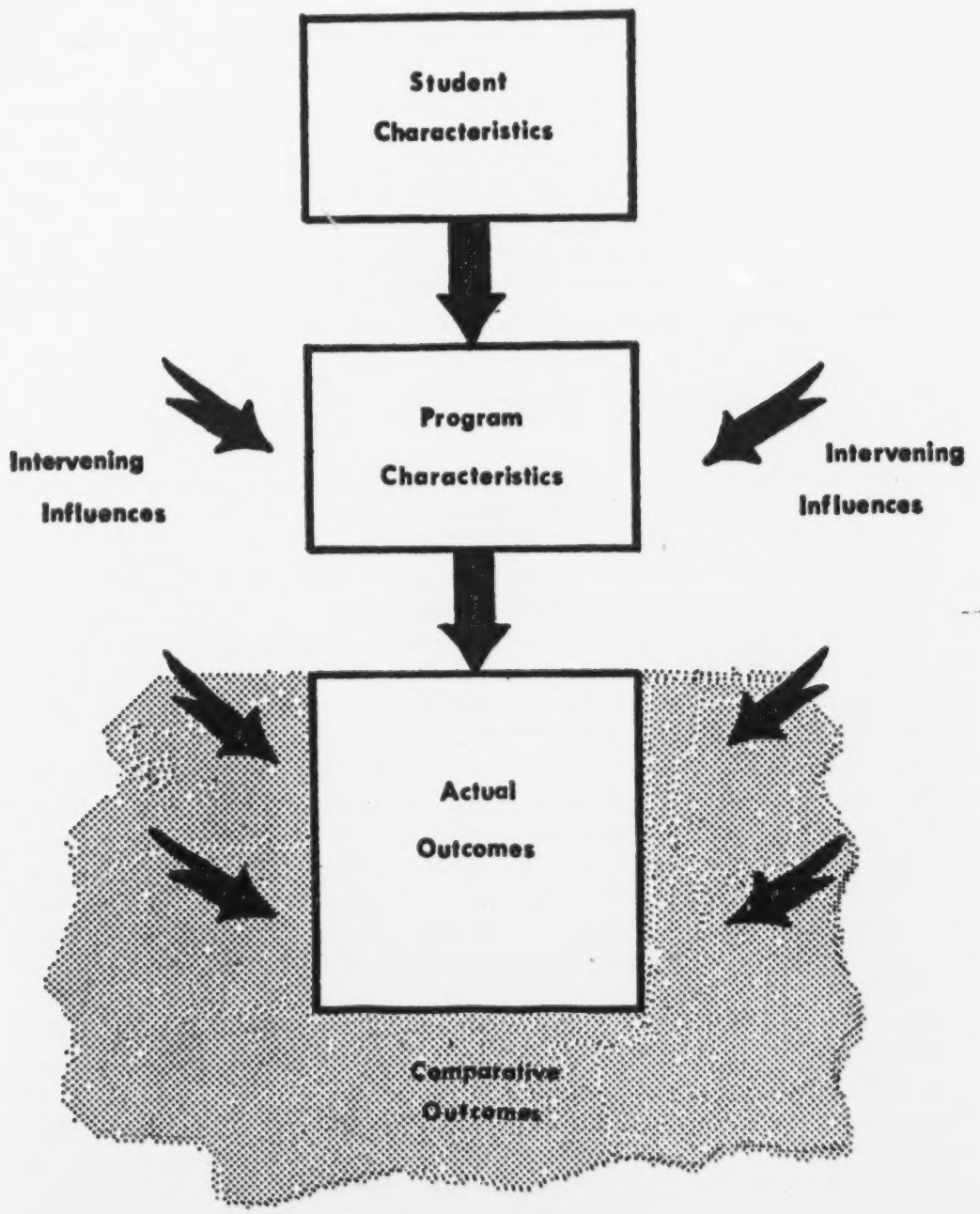

Fig. 1. Major Components of the Evaluative System

provide the comparative standard by which the relative merits of a given program will be judged.

The components of the evaluative system can be utilized to create a more formal definition of program evaluation, as follows:

> Program evaluation is the process of attributing differences between actual and comparative outcomes to program characteristics, under different conditions of student characteristics and other intervening influences, and making a judgment about the value of the program characteristics. The process is conducted for the purpose of making more rational decisions about programs.

Note that the definition has two important qualities. First, evaluation must

be comparative. To report actual outcomes provides only a description of what happened. Evaluation requires making a judgment, which in turn necessitates comparing outcomes with some other set of expected or actual outcomes.

Second, evaluation requires that differences in the outcomes compared must be *attributable* to program characteristics or the interaction of program and student characteristics. Comparing outcomes which do not reflect actual differences in program, but which are due, for example, to differences in students, or to other relevant influences, would be completely misleading.

PROGRAM OUTCOMES (OR EVALUATION CRITERIA)

One of the most critical aspects of program evaluation, and the one which has thus far probably received the least attention, is the identification and measurement of the program outcomes which are to serve as evaluative criteria. Everyone affected by evaluation, and that is all educators, must be concerned with developing as complete an array of relevant, potential outcomes as possible for use by evaluators. What seems needed first is a classification schema, matrix, or topology which will help to identify the scope and boundaries of such an array of outcomes, and to suggest possible criteria for including specific outcomes. One such multi-dimensional classification schema, with some guidelines for filling it with appropriate criteria, is presented below.

Guidelines

First, the criteria by which instructional programs are to be evaluated must be the outcomes—the products—of instruction. Program characteristics cannot be used as evaluative criteria, for, by so doing, we assume, rather than prove, that those characteristics are good. Given the present state of knowledge, the major purpose of evaluation must be to determine which program characteristics actually produce the desired outcomes for a certain group of students. Almost none of our cherished "principles" of vocational education practice have been empirically validated. They have about as much scientific status right now as old wives' tales. Many of them, in time, may prove to be pedagogically sound. But the point is that they remain to be proven. Until they are proven, alleged evaluations based only on the presence or absence of certain program characteristics are acts of faith. In fact, they merely serve to describe the program in terms of variables which we presently think are important.

Second, the matrix of evaluative criteria should include the potential outcomes relevant to each of the different philosophies or value systems under which vocational, or technical, or practical arts programs might be operated. To illustrate, vocational education may be provided primarily to maintain the supply-demand equilibrium of the labor market, or it may be directed mainly toward the development of individuals. Different criteria of success, or at least criteria with different emphases, would be used to reflect accurately each philosophical position. Similarly, manpower training programs can be thought of principally as relieving poverty or as meeting labor shortages and upgrading the labor force. Here too the criteria of success, the expected outcomes, would differ. Reduction of unemployment among the poor would be a key index of relieving poverty, while the reduction of structural unemployment would be a particularly germane index of an upgraded labor

force. The major consideration is that expected outcomes consistent with each of these value systems would be included in the complete evaluative criteria matrix for vocational programs. When two programs of different philosophical bents are compared, both sets of pertinent criteria must be used as the basis for comparison in order to provide a clear picture of each program's relative strengths and weaknesses. In fact, no program should be evaluated only in terms of the objectives stated by its developers or operators. The question is always what does the program gain and what does it give up. We must observe whether the hoped for outcome occurs, while at the same time making certain that damaging effects are not offsetting gains. These "side-effects" of programs are too often completely ignored, and at great potential peril to the long-range interests of students and educational institutions. Careful examination of the variables related to the optimized program characteristics and to the anticipated outcomes will help reveal potential side-effects.

Third, the expected outcomes of educational programs can be stated at several levels of specificity. At the most general (macro) level, expected outcomes are broadly stated to express directly a philosophical position. At more specific levels, outcomes can be stated as measurable indices or as operational definitions of the more generalized statements of expected outcomes. At their most specific (micro) level, outcomes can take the form of items on an achievement test, a questionnaire, a rating scale, a personality inventory, and so on, which samples behavior under various stimulus and response conditions. Obviously, each succeeding level of greater specificity must represent a valid sample of the more generally stated ends if measures are to accurately reflect basic intents. To illustrate, a general outcome or objective for a vocational program might be to upgrade the quality of the labor force. One, more specific, expression of that hoped-for outcome would be to increase the ability of students in preparatory programs to use their skills in a wide variety of situations. A measurable index of that ability could be the time needed to retrain for a related occupation, to some specified level of performance. An operational definition of that index might be the time it will take to train secretarial graduates of a given ability level from "X" vocational school to type 130 three-inch wide justified lines per hour on a Varityper with no more than five errors. The actual performance test and the conditions under which it is given and scored constitute one example of the micro level of the expected outcome. Of course, many additional indices, operational definitions, and items would have to be sampled in order to properly measure whether or not the quality of the labor force had been upgraded.

For the purpose of program operators and evaluators, the profession should suggest outcomes expected at both the most general, philosophical level, and at the level of measurable indices. At the present time, this would require the development of at least two matrices for vocational and technical programs, and two for practical arts programs (see Figure 2).

Fourth, indices of program outcomes should not only be consistent with philosophical positions, but they should also be sensitive to variations in program characteristics. As far as possible, they should logically and directly reflect differences in programs, while remaining relatively independent of student characteristics and other intervening, e.g., economic, conditions. The function of program evaluation is to attribute differences between outcomes to program characteristics. Gross indices, or those which are influenced primarily by non-educational variables, will make the task of evaluation extremely difficult, if not impossible, by hiding, rather than highlighting, the effects of program characteristics.

Fifth, it would greatly facilitate weighing the relative merits of programs if it were possible to assign monetary values to program outcomes. Summated or per student dollar benefits of different programs could then be directly compared. If it were also possible to measure the dollar cost of providing each program, an index of investment return per program could be obtained. These cost-benefit ratios would be most useful measures for comparing the relative efficiency of various programs. Unfortunately, economic analysis has greater potential than it has current utility. A major problem is in placing valid monetary values on all relevant outcomes. This restricts the usefulness of economic analysis in that only a limited number of outcomes can be employed in the evaluation. One alternative might be to agree to assign dollar values to various outcomes in a manner consistent with some social philosophy. Such an arbitrary technique could serve to make explicit the implicit weightings individuals inevitably give to different kinds of outcomes when judging overall program effectiveness. It seems important to reiterate, though, that whether economic analyses of programs are limited or complete, they yield data which must be interpreted in terms of given student characteristics and other intervening influences on program outcomes. We must always interpret program cost and benefit in terms of Johnny, and not Johnny in terms of cost and benefit.

A Multi-Dimensional Classification Schema

The classification system for expected outcomes should be multi-dimensional. One likely prospect is a three-dimensional schema consisting of a time axis, a target axis, and a type axis. Figure 2, "Matrix of Expected Outcomes," depicts this possibility.

One axis of the matrix, representing elapsed time since the educational experience was provided, is needed to indicate when it is feasible to measure each expected outcome. One category, for example, might be "immediate," to include all the expected outcomes which can reasonably be measured while the student is still in the educational program. Measures of achievement, both in the program and in other courses taken concurrently, drop-out rate, etc. belong in the "immediate" category. A second category, "intermediate," could include outcomes expected to be manifest within some arbitrary but convenient period, such as four years after leaving the program. Grades earned in subsequent college or vocational and technical programs, time required to find initial employment, the time needed to become satisfactorily productive on the job, etc. are illustrative of "intermediate" outcomes. A third time category might be "long-range," and include those outcomes which are feasible to measure more than four years after the educational experience. Voting record, income, and the salience and centrality of work-oriented values are among potential "long-range" outcomes. Some outcomes, of course, can be measured in two, or even all three, of the time categories.

It should be recognized that as elapsed time increases so does the difficulty of data collection and the number of intervening influences, thus complicating the problem of attributing outcomes to the educational experience. Consequently, it seems preferable to devote primary attention at this time to the measurement of "immediate" and "intermediate" outcomes.

One very interesting basic issue, which deserves careful study and discussion, shall be noted only in passing. To what extent should educational programs take responsibility for the overt behavior of students (or former students) in out-of-school situations in which many of the immediate factors influencing

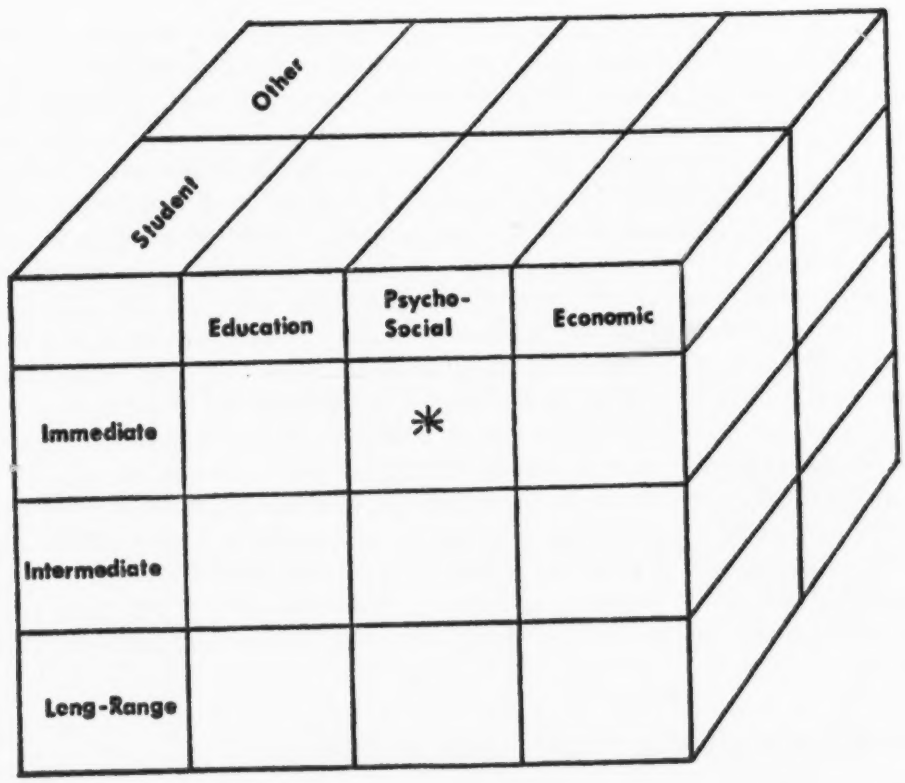

Fig. 2. Matrix of Expected Outcomes (One Matrix for Each Level of Specificity)

that behavior are beyond the control of the school program? Conversely, to what extent do we discharge our responsibility to the individual and society by showing the student's capability to act by his performance under controlled conditions? The answers to these questions will greatly influence our perception of appropriate program outcomes.

A second axis, for target, distinguishes between expected student (and former student) outcomes, and indirect, secondary, or feedback outcomes anticipated in other people, agencies or institutions. We might call these two categories "student" and "other." The "student" category would logically include an expectation of an "immediate" increase in vocational maturity among program enrollees; the "other" category could contain an index of "immediate" improvement in parental attitudes toward the school. Some additional illustrative outcomes in the "other" category are effects upon teacher morale, educational administrative patterns, and community support for education. The "other" category might also include indices of the program's impact upon firms, occupations, industries, and even upon the community, the economy, and society as a whole.

A third axis is needed to distinguish among types of expected outcomes. One possible distinction is among educational, psycho-social, and economic criteria. Educational outcomes are illustrated by the cognitive and psychomotor abilities acquired in the program, and by the numbers of persons receiving training in relation to those needing or requesting it. Reduction in the delinquency rate of students, and the attitudes of prospective employers

or of other educational institutions toward the program, are examples of psycho-social outcomes. Economic criteria include earnings of former students and labor cost per unit of production.

It may be postulated that qualitative and quantitative outcomes also deserve separate sub-categories on one of the three suggested axes. However, rather than complicate the classification schema further, it seems preferable to include both qualitative and quantitative outcomes, where applicable, in every cell.

The benefits of developing a complete array of relevant, potential outcomes for use by evaluators can be summarized as follows:

a. The process of program development and revision will require continuous clarification of extant philosophical positions and examination of their internal consistency. The array will facilitate this activity, and will also point to the kinds of new indices and measurement devices that need to be created.
b. The classification schema will provide the basis for cumulating and organizing the results of evaluative studies and for identifying directions for further efforts.
c. A less obvious but equally significant benefit of an array of expected outcomes is to place each evaluative study in proper perspective. That is, all evaluation studies will undoubtedly be partial in terms of the limited number of outcomes actually measured. Program operators will become more aware of these criterion limitations, and thus be better able to make more rational policy and administrative decisions based upon the results of given studies.

PROGRAM CHARACTERISTICS

Program characteristics may be thought of as the sub-systems, such as the teacher, content and content organization, methods and techniques of instruction, facilities, etc., which provide the student with relevant, planned instructional experiences. The effects of all or one of these sub-systems may be the subject for evaluation—the independent variables—in any given style.

A careful description of program characteristics is prerequisite to evaluation. Many studies fail at this very basic level. The characteristics to be described need both careful definition and ratings using acceptable absolute scales should be employed, so that greater interrater reliability is possible, and so that the same characteristics can be meaningfully compared across programs.

Because evaluation is a time-consuming activity, and each study can only provide data on a limited number of outcomes, it seems important that the attention of research-evaluators be focused initially upon programs with basically different practical and theoretical characteristics (sub-systems). In vocational education, for example, we need to turn our attention first to comparisons of (a) vocational, general, and college preparatory programs, (b) in-school educational efforts, cooperative programs, and on-the-job training, (c) programs that include a large component of "general" education content, and those that do not, (d) programs which focus upon relatively specific occupational goals, and those that prepare each student for a broad range of occupations, (e) programs that integrate occupational content inductively rather than deductively, (f) programs that provide for individual differences by varying content and method, and those that do not, (g) programs which utilize enrollment time as an individualized variable, and those that do not, (h) programs which incorporate information about the psychological and

sociological elements of the occupational environment, and those that do not, and so forth.

Each of these comparisons should be made using a variety of content drawn from different kinds of occupations until we learn if and how subject matter differences interact with other kinds of program characteristics to affect program outcomes. The nature of the content being taught (learning task) must, until we learn otherwise, be considered an important program characteristic in the evaluative system. Similarly, the potential variation in outcomes due to the effect of individual teacher differences seems to be great enough to require that some attempt be made to account for their influence.

Technology is fact making it feasible to individualize instruction. Consequently, the days when researchers were forced to try to find the one best method or curriculum for all students are past. The problem now is to identify the configurations of student characteristics and program characteristics (including content) which produce desirable outcomes with maximum efficiency. It is therefore important that comparisons of program characteristics be made so that program outcomes can also be related to carefully described differences in student characteristics.

As previously noted, it would be desirable to interpret all program outcomes in terms of their dollar value. But it is *essential* to meaningful evaluation that the costs of providing each program be measured. Costs are the inputs, in resources and time, utilized by the program to produce outcomes, under certain conditions of students and influences. So long as resources for education remain limited, any realistic evaluation must be concerned with program efficiency, as well as effectiveness. Cost per unit outcome (even if unit outcome is not in dollars) is such a measure of efficiency. And, as long as final outcomes are comparable, the relative efficiency of different programs can be ascertained by using this measure. In computing program cost, social ("other") as well as "individual" costs must be determined, just as social ("other") and "individual" program outcomes may be considered benefits. Unfortunately, most of our present reporting systems are not producing the kind and quality of data needed to make complete, accurate program cost estimates. Most schools are thereby failing in their responsibility to the public for accountability.

TWO ROLES OF PROGRAM EVALUATION

As many authors have pointed out, evaluations can be conducted by different people for many different kinds of reasons. There are, however, two primary roles played by evaluation in the process of answering two different sets of basic questions.

First, what Scriven (1967) has called the "formative" role of evaluation is an attempt to aid in the process of developing or improving a program. It answers questions like: How well is the program accomplishing what it set out to do? How can it be improved? How can program efficiency be maximized in terms of accepted program goals? Second, Scriven's "summative" role of evaluation is an attempt to estimate an operational program's overall effectiveness; it provides a basis for choosing among programs. It answers questions like: Which of these two programs is better for my purposes? What would I gain and lose by adopting the new curriculum?

The differences between these two kinds of questions engender some important distinctions in the criteria employed by formative and summative evaluations. Formative evaluation derives its comparative base from within the same program. Comparisons may be between actual and expected out-

comes, present and past outcomes, or between concurrent sets of outcomes derived from manipulating certain program characteristics. In the case of developmental projects, immediate outcomes are typically used; for improving established programs, intermediate outcomes are sometimes employed as criteria. Parts of the program or the total program may be the subject for evaluation.

On the other hand, summative evaluation can best be accomplished by comparing the outcomes of two or more programs. Since the programs may have different intended outputs, the evaluative criteria employed must include all of the relevant expected outcomes. The outcomes used are usually intermediate, since they frequently provide a more meaningful basis for comparing different programs than do immediate outcomes.

EVALUATION AS A PART OF THE EDUCATIONAL CHANGE PROCESS

Another way to conceptualize the roles of evaluation is by considering its functions in the educational change process.

It may well be too soon to assess the extent to which recent federal assistance for research-related activities has resulted in substantive changes in educational programs, but it is already obvious that one extremely significant outcome of that intervention has been a concern for systemizing and facilitating the very process of educational change. Where formerly it was accepted that change would be slow, erratic, and uncoordinated, the availability of relatively large amounts of money for research and development has instigated the study of the educational change process; it has become necessary to decide how to invest those funds in order to insure rapid, qualitative improvements in education.

Figure 3, "Educational Change Model," illustrates one version of a generalized abstraction of the change process. It draws heavily upon the prior work of Guba and Clark (1965). Note the four places in the model that call for evaluative efforts.

The first place is in the process of inventing and engineering innovative educational products. Development must provide for formative evaluation as an integral part of the process; developers must know whether their new instructional program or device is performing according to expectations; they must diagnose early results in order to improve upon efficiency and effectiveness. Similarly, developers should be responsible for beginning to accumulate data for a summative evaluation so that program operators can eventually make more rational decisions about the comparative value of the new versus the standard program or device.

During the diffusion stage, in which innovative programs or devices are being demonstrated (on the strength of the results of prior formative evaluations and preliminary summative evaluation data), it is inevitable that formative evaluation will be continued. It should also be required of developmental projects that additional data, including immediate and intermediate outcomes, be collected during the demonstration phase for use in making between program comparisons.

When individual program operators or states elect to adopt a new program or device, presumably on the strength of a pertinent summative evaluation, there is typically a stage of adaptation and modification which calls for further formative evaluation.

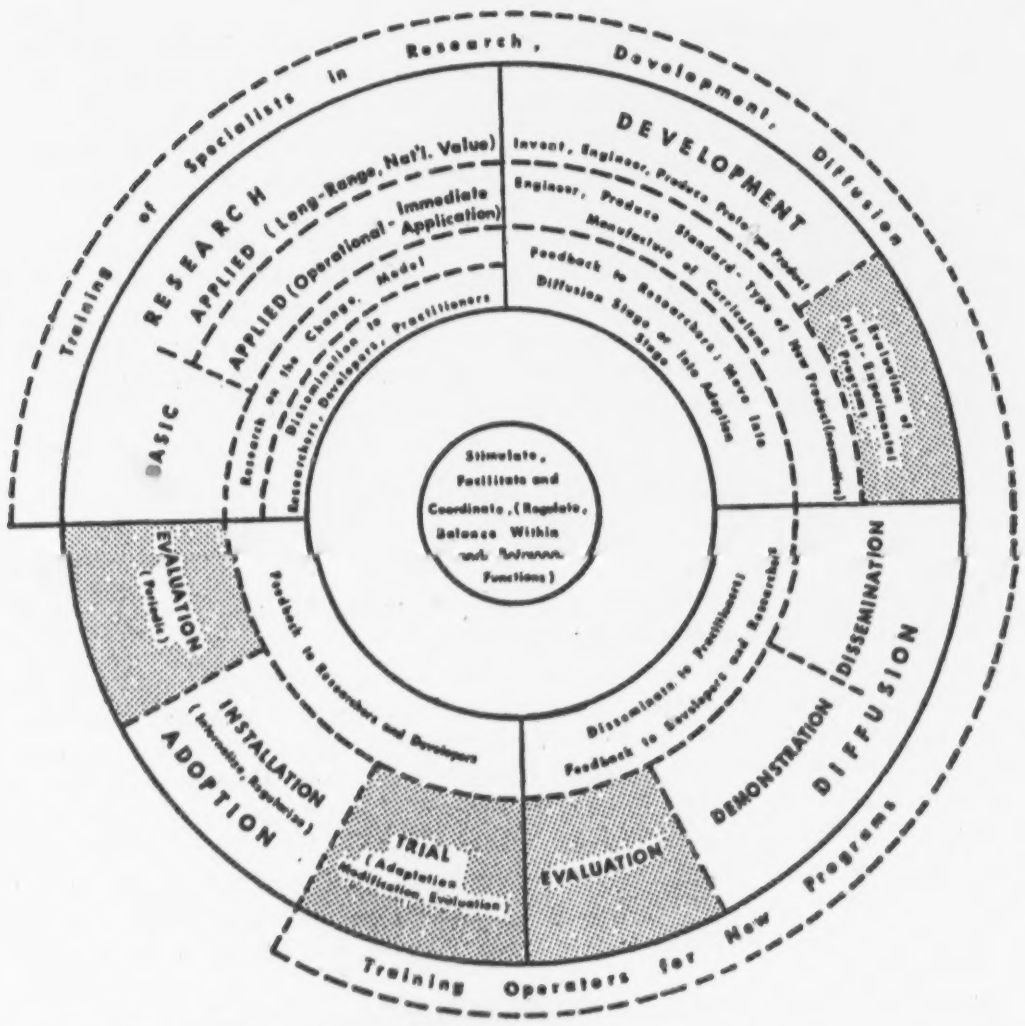

Fig. 3. Educational Change Model (Read Clockwise)

Finally, all operational programs need periodic evaluation. These assessments should be formative, to improve upon the existing basic program, and summative, to test its efficiency and effectiveness against other programs, including the new products of later developmental projects.

SOME RESEARCH APPROACHES TO EVALUATION

Formative Evaluation

Formative evaluations are conducted using a variety of research approaches: (a) experts study the content of authenticity, and check the logical consistency among general purposes, students, content, method, and performance measures; (b) feedback from program developers, evaluators, teachers, students, former students, etc., based upon direct or indirect observation of and/or participation in the program, provides clues for making program changes; (c) pre- and post-tests of a diagnostic nature supply more objective information for making potential improvements in the program; (d) experiments, using program characteristics as independent variables, are sometimes employed; (e) developmental, reproduction, and operational costs should be (although too frequently are not) estimated.

Unfortunately, it is usually difficult to know when to be satisfied with the results of a formative evaluation—when to stop revising content and method and striving for greater efficiency—and when to begin revising measurement devices and expected outcomes. Even if performance objectives are stated very specifically in advance of program development, our present lack of sophistification in setting and stating objectives would lead us either to question their validity or to continue striving for greater efficiency in their attainment.

There is an obvious need for summative evaluation wihch provides for making objective comparisons among programs. A variety of research approaches has been employed in an attempt to satisfy this need, including evaluation by expert- or self-judgments, follow-ups, experiments, quasi-experiments, and regressional analyses. (Many of these approaches are also useful in formative evaluation studies.)

Expert and Self-Evaluation

The judgment of program operators, developers, consultants, etc., can, as previously indicated, be very helpful in formative evaluation by suggesting revisions in program characteristics. Judgments (ratings) of program characteristics can also be helpful in summative evaluation by providing "sensitive" descriptions of the extant programs. But it is thoroughly misleading, and as inappropriate use of expertise, to utilize judgments about program characteristics as criteria in summative evaluation. Since programs must be developed, operated, and perhaps even accredited right now, it is necessary that judgments be made about the characteristics which constitute "good" programs. We must not, however, delude ourselves into believing that these judgments are necessarily correct; care must be taken not to compound possible judgment errors. Prior judgments must be put to the test of reality whenever programs undergo summative evaluations.

Follow-Ups

The follow-up is a procedure which gathers data about former students. Some times the procedure is employed to secure the opinions of former students about the program. If the investigator is willing to accept these judgments, they may be used in formative evaluation to suggest program revisions. More frequently, the follow-up is used to collect data about the status of former students to serve as program outcome criteria. For instance, information about their work history may be obtained. These data can provide evidence for use in both formative and summative evaluations, provided that the potential differential effects on outcomes of intervening variables and student characteristics are taken into account. For summative evaluations the adjusted data about former students must be compared with equivalent adjusted data from some alternative program. Too many studies have been reported in which a high placement rate (as one program outcome) is assumed to be valid evidence of a good vocational program, without bothering to compare that rate with some alternative program's placement rate, or without taking into consideration possible differences between programs in student aptitudes and in labor demand in the geographical areas concerned.

Because follow-ups focus upon former students, they should not be the only means used for collecting data on program outcomes. The impact of programs on the school, the industry, the community, etc., are liable to escape attention unless other data collection procedures are also employed.

Experiments

The experiment is by far the most efficient design for the conduct of summative evaluations, and should be used whenever possible. It controls for differences in student selection, and many other kinds of extraneous, intervening variables (history). Some natural opportunities for utilizing true experiments do present themselves. To illustrate, inadequacy of resources, or the need for "tooling-up" time, may necessitate a staged introduction of a new program into a school system, thus providing an opportunity for control and experimental groups of schools or classes. In other situations, more volunteers than can be handled may be recruited, and random assignment provides a desirable social and scientific means for forming treatment groups. Teachers might also be asked to nominate matched pairs of students, each of whom is likely to want to cooperate in a new program, and these pairs can then be randomly divided into experimental and control groups before actually requesting the experimental group to volunteer.

Interrupted Time Series

It is most frequently the case, however, that true experimental designs are not feasible to summative evaluation studies. In these situations, the interrupted time series (Campbell & Stanley, 1963; Campbell, 1967), a quasi-experimental design, should prove extremely useful. The essence of the time series is the periodic measurement of some unit and then the introduction of an experimental change in the unit, the results of which become evident by a discontinuity in subsequent periodic measures. Figure 4 illustrates four examples of this approach.

Assume that an outcome (or even cost per unit of some outcome) of a one-year vocational program, such as percent of immediate placement in a related occupation, is measured each year for succeeding classes so that measures are available for 1966, 1967, and 1968. Then, in 1968-69, a major revision is introduced into the program. Percent of immediate placement in a related occupation is again measured for the 1969, 19670, and 1971 classes.

Was the program revision successful? If time series A or B, in Figure 4, occurred, then it appears that the revision had no effect, but, if examples C or D were found, it would appear that the new program had a very desirable effect on the measured outcome. In fact, if the same criterion measure was used in all years, the only two plausible alternative explanations of the changes in time series C and D, other than the revised program, would be abrupt differences in student characteristics and/or other historical, e.g., socio-economic, influences after 1968. If data is available for all years on students and on germane historical conditions, these alternative explanations of the jump in outcome can either be eliminated, or regressional analyses can be used to adjust the outcomes for their differences. Should comparable data also be available for the same years from other programs in the same labor market area, it is possible to begin to use cross-program comparisons (multiple time-series) to eliminate some historical conditions as possible sources of internal invalidity.

It is obvious that, to be considered significant, the change in measured outcome should exceed the fluctuation normally expected in the time series. The logical test of the hypotheses of no change would therefore be between some extrapolated-expected outcome of the time series which preceded the experimental intrusion and the actual outcome after the experimental treatment. If continuous improvement, as in example D, is hypothesized, Campbell (1967)

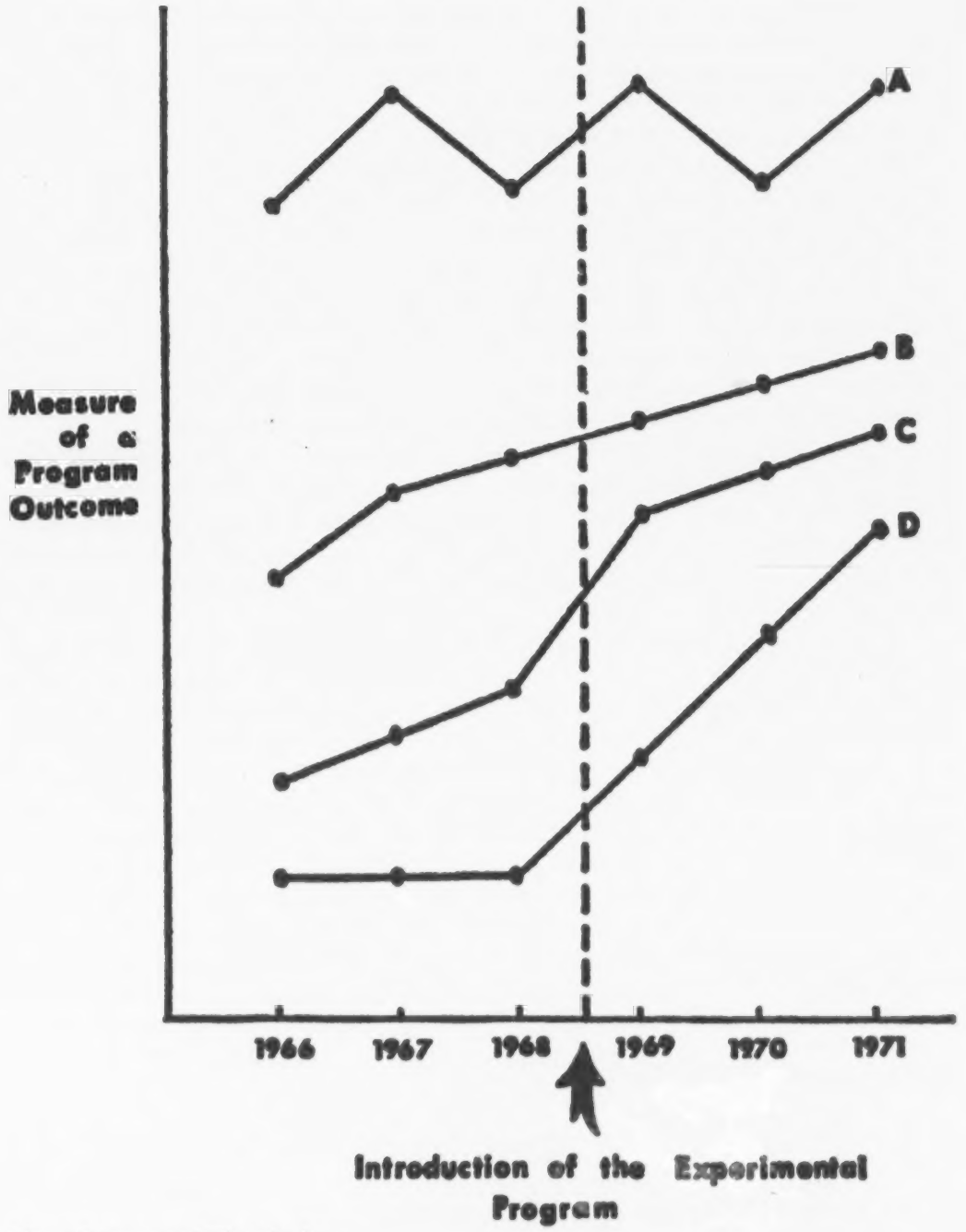

Fig. 4. Interrupted Time Series

suggests that the appropriate test is between the slopes of the time series before and after the experimental treatment (in addition to a test of difference in expected and actual intercepts). Glass and Maguire (1968) have recently investigated the applications of certain statistical models for analyzing the change in level of performance.

Regressional Analysis

Regressional analysis has recently begun to be employed in formative and summative evaluation studies as a statistical technique for accounting for variation in program outcomes that are related to differences in student characteristics and other intervening variables.

The evaluator proceeds by gathering data on outcomes, student characteristics, and intervening variables from the programs to be evaluated. Assuming two programs are being compared in a summative evaluation, a full multiple linear equation of the general form shown as 1) in Figure 5 is developed for the combined programs, and R_f^2 is determined.

Then, as also shown in Figure 5, P_1 and P_2 are deliberately made equal to O and a new R_r^2 is computed from the restricted equation. The significance of the difference between R_f^2 and R_r^2 can be tested by the F statistic. A significant reduction indicates that program membership does account for a significant amount of the variation in program outcome, independent of the student characteristics and intervening variables used in the equation. It means, in effect, that after accounting for differences in program outcomes due to variations in certain student characteristics and intervening variables the kind of program provided still has a significant relationship to outcomes.

Other general equations are available to test for interactions between program and student characteristics and between program and other influences. In addition, second and third degree polynomials can describe curvilinear functions. Canonical correlations can relate combinations of independent variables to combinations of dependent program outcome variables, all in one equation.

Regressional analysis is therefore a most useful, almost indispensable tool in evaluation. There are, however, certain limitations of the technique which should be made explicit. First, a fairly large sample of students, which increases with the number of independent variables used, is needed to obtain reliable results. Second, unless the analysis can be combined with a true experimental design, or unless $R_f^2 = 1$, the relationships yielded by the equations are *not* necessarily casual. The relationships found are based entirely upon the variables used in the equation; the use of additional variables could change the relationships found. Third, the relationships revealed are peculiar to the sample used, or to other groups of students of which the sample is representative. The sample, for example, might happen to consist of students with quite similar IQ scores. IQ would then probably not reveal a high relationship to achievement as a program outcome because of its restricted range. In another more heterogeneous sample, IQ could, and probably would, be found to have a strong relationship to achievement.

---

1) $Y' = a_1 + b_1P_1 + b_2P_2 + b_3S + b_4I$, which yields R_f^2

where Y' is a predicted program outcome

a_1 is a constant (computed)

P_1 is membership or not in one of the two programs

P_2 is membership or not in the second program

S is values from student characteristic(s)

I is values from intervening variable(s)

b_1-b_4 are partial correlation coefficients (computed)

R_f^2 is the percent of variation in Y explained by the full equation

2) Setting $P_1 = P_2 = 0$:

$Y' = a_2 + b_5S + b_6I$, which yields R_r^2

Fig. 5. Regression Analysis Model

IMPLICATIONS FOR ACTION

It is quite urgent that states begin to plan for the periodic collection of data which are necessary to evaluate existing programs. This paper has provided a rationale and guidelines for planning and conducting those evaluations. It has recommended that the data to be collected include measures of (a) student characteristics, (b) program characteristics and program costs, (c) intervening variables, and (d) program outcomes. But it remains for states to decide upon the specific variables to be utilized, the means for measuring them, and the techniques for collecting, storing, and retrieving the measures.

Judgments about the potential fruitfulness of relevant variables and their measurement must be made, based upon presently available knowledge and expert opinion. Even though some decisions may later prove to be incorrect, the "evaluative system" should be established just as soon as it is possible to conduct careful reviews and to make considered judgments. Counsel from a wide variety of specialists, within and outside of education, must be an integral part of the decision-making process. Researchers can then assume the long-term task of seeking new knowledge to improve the system.

And the federal government has a pivotal role to play. It has a responsibility to stimulate activity, to facilitate communication among states, and to provide consultation throughout the development of the system. Finally, it must assure that some reasonable minimum amount of information to be collected for descriptive and evaluative purposes is comparable among the states.

Selected Bibliography

Campbell, D. T. Administrative experimentation, institutional records, and nonreactive measures. In J. C. Stanley (Ed.) *Improving experimental design and statistical analysis.* Seventh Annual Phi Delta Kappa Symposium on Educational Research. Chicago: Rand McNally & Co., 1967, 257-291.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.) *Handbook of research on teaching.* A project of the American Education Research Association. Chicago: Rand McNally & Co., 1963, 171-246.

Glass, G. V., & Maguire, T. O. Analysis of time-series quasi-experiments. Final report of Project No. 6-8329 n.p. Washington, D. C.: Office of Education, U. S. Department of Health, Education, and Welfare, March 24, 1968, 101.

Guba, E., & Clark, D. An examination of potential change roles in education. Paper presented at the N. E. A. Conference on Innovation in Planning School Curricula, Airlehouse, Virginia, October 2-4, 1965.

Oettinger, A. C. The myths of educational technology. *Saturday Review.* May 18, 1968, 76-77, 91.

Stake, R. E. The countenance of educational evaluation. *Teachers College Record,* 1967, *68,* 523-540.

EVALUATION OF TRAINING[1]

Donald L. Kirkpatrick, University of Wisconsin, Milwaukee

An effective training director will make an effort to evaluate all of his training activities. The success of these efforts depends to a large extent on a clear understanding of just what "evaluation" means. This chapter will attempt to accomplish two objectives: (1) to clarify the meaning of evaluation, and (2) to suggest techniques for conducting the evaluation.

These objectives will be related to "in-house" classroom programs, one of the most common forms of training. Many of the principles and procedures can be applied to all kinds of training activities such as performance review, participation in outside programs, programmed instruction, and the reading of selected books.

The following quotation from Daniel M. Goodacre III[1] is most appropriate as an introduction:

> Managers, needless to say, expect their manufacturing and sales departments to yield a good return and will go to great lengths to find out whether they have done so. When it comes to training, however, they may expect the return—but rarely do they make a like effort to measure the actual results. Fortunately for those in charge of training programs, this philanthropic attitude has come to be taken for granted. There is certainly no guarantee, however, that it will continue, and training directors might be well-advised to take the initiative and evaluate their programs before the day of reckoning arrives.

EVALUATION CLARIFIED

Nearly everyone would agree that a definition of evaluation would be "to determine the effectiveness of a training program." But this has little meaning until we answer the question, "In terms of what?" We know that evaluation is

---

[1] Published in R. L. Craig and L. R. Bittel (Eds.) *Training and Development Handbook*. New York: McGraw-Hill Book Company, 1967, and reprinted with the permission of the author and publisher.

needed in order to improve future programs and to eliminate those programs that are ineffective. The problem is how to begin.

Evaluation changes from a complicated elusive generality into clear and achievable goals if we break it down into logical steps. These steps can be defined as follows:

Step 1—*Reaction.* How well did the conferees like the program?

Step 2—*Learning.* What principles, facts and techniques were learned?

Step 3—*Behavior.* What changes in job behavior resulted from the program?

Step 4—*Results.* What were the tangible results of the program in terms of reduced cost, improved quality, improved quantity, etc?

With this clarification of the meaning of evaluation, training men can now begin to pinpoint their efforts at evaluation. They better realize what they are doing and they recognize the limited interpretations and conclusions that can be drawn from their findings. As they become more experienced and sophisticated in evaluation design and procedures, they slowly begin to obtain more meaningful results on which future training can be based.

These four steps will now be defined in detail with examples and suggested guideposts. It is important to stress that the described *procedure and techniques* can be used in almost any organization. It is also important to stress that the *results* from one organization cannot be used in another organization. Obviously, there are many factors that would influence the results. These variables include the group, the conference leader, and the approach to the subject.

STEP 1—REACTION

Reaction may best be defined as how well the trainees liked a particular training program. Evaluating in terms of reaction is the same as measuring the feelings of the conferees. It is important to emphasize that it does not include a measurement of any learning that takes place. Because reaction is so easy to measure, many training directors do it.

Guides for Evaluating Reaction

1. Determine what you want to find out.
2. Use a written comment sheet covering those items determined in step one above.
3. Design the form so that the reactions can be tabulated and quantified.
4. Obtain honest reactions by making the forms ananymous.
5. Allow the conferees to write in additional comments not covered by the questions that were designed to be tabulated and quantified.

The comment sheet shown in Fig. 1 was used to measure reaction at an ASTD summer institute that was planned and coordinated by the staff of the Management Institute of the University of Wisconsin.

Those who planned this ASTD program were interested in reactions to

subject, technique (lecture versus discussion), and the performance of the conference leader. Therefore, the form was designed accordingly. So the reactions could be readily tabulated and quantified, the conferees were asked to place a check in the appropriate spaces.

In question 3 concerning the leader, it was felt that a more meaningful rating would be given the leader if the conferees considered items A through G before checking the overall rating. This question was designed to prevent a conference leader's personality from dominating group reaction.

Question 4 allowed the conferees to suggest any improvements that came to mind. The optional signature was used so that follow-up discussions with conferees could be done. In this ASTD summer institute, about half of the conferees signed their names. With this type of group, the optional signature did not affect the honesty of their answers, in all probability. It is strongly suggested that unsigned sheets be used in most in-house meetings, however.

This ASTD reaction sheet was used at the conclusion of every session in the institute program. Therefore, the conferees rated each conference leader for his contribution to the program. In many internal training programs, a series of meetings will be held and the reaction sheet will not be used until the end of the last session. This is especially true when one conference leader conducts the entire program. In this case, a comment sheet like the ASTD one might be adapted to the situation and modifications made.

A very practical suggestion came from Richard Johnson of Con-Gas Service. He suggested that the comment sheets be given to the enrollees before the program is over so that the suggestions can be used in improving the last section of the training program. For example, where a training program consists of a series of nine sessions, the comment sheet should be given to conferees at the end of the third session. Their comments and suggestions should be taken into consideration to make the last six sessions more effective.

How to Supplement the Evaluation of the Conferees

It has been emphasized that the form should be designed so that tabulations can be readily made. In this writer's opinion, too many comment sheets are still being used in which the conferees are asked to write in their answers to questions. Using a form of this kind, it becomes very difficult to summarize comments and to determine patterns of reaction.

At the Management Institute of the University of Wisconsin, every session is evaluated in terms of the reactions of the conferees. This has been done for more than 15 years. Occasionally the coordinator of the program felt that the group reaction was not a fair evaluation of the effectiveness of the program. Sometimes the staff men felt that the conference leader's personality made such an impression on the group that he received a very high rating. In other sessions, the coordinator felt that the conference leader received a low rating because he did not have a dynamic personality. Therefore, the Management Institute adopted a procedure by which every conference leader is rated by the coordinator as well as by the group. A form similar to Fig. 2 can be used.

This procedure in which the coordinator of the program also evaluates each conference leader was also used in an ASTD summer institute. It was found that a coordinator's rating was usually close to the group's rating, but in some instances it varied considerably. In selecting and orienting future conference leaders for ASTD institutes, both of the evaluations are taken into consideration.

ASTD Institute

Leader_____ Subject_____

Date_____

1. Was the Subject Pertinent to Your Needs and Interests?

☐ No ☐ To Some Extent ☐ Very Much So

2. How Was the Ratio of Lecture to Discussion?

☐ Too Much Lecture ☐ O.K. ☐ Too Much Discussion

3. Rate the Leader on the Following:

	Excellent	Very Good	Good	Fair	Poor
A. How well did he state objectives?					
B. How well did he keep the session alive and interesting?					
C. How well did he use the blackboard, charts, and other aids?					
D. How well did he summarize during the session?					
E. How well did he maintain a friendly and helpful manner?					
F. How well did he illustrate and clarify the points?					
G. How was his summary at the close of the session?					

What is Your Overall Rating of the Leader?

☐ Excellent ☐ Very Good ☐ Good ☐ Fair ☐ Poor

4. What Would Have Made the Session More Effective?

_____

_____

_____

_____

_____

Signature (optional)

Fig. 1. Rating Chart

It is suggested that the training director in each company consider this approach. A trained observer such as the training director or another qualified person would fill out an evaluation form independent of the group's reactions. A comparison of the two would give the best indication of the effectiveness of the program.

Coordinator's Rating of Leader

Date_____

Rating_____ Name of Leader_____ Subject_____

	Very Much So	To Some Extent	No
A. PREPARATION 1. Did he prepare for the meeting?			
2. Was his preparation geared to the group?			
B. CONDUCTING 1. Did he read his material?			
2. Did he hold the interest of the group?			
3. Was he enthusiastic-dynamic?			
4. Did he use visual aids?			
5. Did he present his material clearly?			
6. Did he help the group apply the material?			
7. Did he adequately cover the subject?			
8. Did he summarize during conference and at end?			
9. Did he involve the group?			

C. CONSTRUCTIVE COMMENTS

1. What would you suggest to improve future sessions?

_____

_____

D. POTENTIAL

1. With proper coaching what would be the highest rating he could achieve? _____

E. ADDITIONAL COMMENTS

_____

_____

Fig. 2. Leader Rating Sheet

Measuring Reactions to Outside Training Programs

The forms and suggestions that have been described above will apply best to an internal training program. Since many companies send their management people to outside training programs at universities, American Management Association, National Industrial Conference Board, etc., it is suggested that the reaction of each person attending such a program be measured. Lowell

Results of University of Wisconsin
Management Institute Program Questionnaire

Key	Programs
A	Modern Leadership for Middle Management
B	Supervisors' Leadership in Cost Control
C	Developing Supervisory Skills
D	Human Relations for Foremen & Supervisors
E	Leadership and Growth
F	Creative Thinking for Supervisors
G	Human Relations for New Foremen
T	Totals

	A	B	C	D	E	F	G	T
Quetionnaires returned:	3	3	5	11	5	1	1	29
Responses:								
1. I thought the program was:								
A. Very well organized and helpful	3	3	5	11	5	1	1	29
B. It was of some value								
C. It was poorly organized and a waste of time								
2. In reference to the subject content:								
A. It was all theory and of little practical value								
B. It was both theoretical and practical	3	2	2	3	1			11
C. It was very practical and useful	0	1	3	9	4	1	1	19
3. Concerning the quality of the instruction:								
A. The instruction was excellent	2	3	4	11	4	1	1	26
B. The instruction was average			1		1			2
C. The instruction was of poor quality								

Fig. 3. Oscar Mayer & Co. Evaluation Form

Reed, former training director of the Oscar Mayer & Company of Madison, Wisconsin, used the form in Fig. 3 for evaluating the reaction to the University of Wisconsin Management Institute program.

In this situation, Oscar Mayer & Company was not interested in the reaction to specific leaders. They were interested in reaction to the overall program to determine whether or not to send other foremen and supervisors. In other words, this particular questionnaire was designed to fit the need of Oscar Mayer & Company.

Another company used the form in Fig. 4 to evaluate the reaction of their managers who attend an outside program.

Conclusions about Reaction

The first step in the evaluation process is to measure the reactions to training programs. It is important to determine how people feel about the programs they attend. Decisions by top management are frequently made on the basis of one or two comments they receive from people who have attended. A supervisory training program may be canceled because one superintendent told the plant manager that "this program is for the birds."

Reaction to Supervisory Institute
by Foremen and Supervisors Who Have Participated

IN GENERAL

1. How worthwhile was the Institute for you?
   ____ Very worthwhile
   ____ Fairly worthwhile
   ____ Not very worthwhile
   ____ A waste of time

2. The Institute had:
   ____ Too much theory and not enough of the practical
   ____ Too much of the practical and not enough theory
   ____ About the right combination of theory and practice

HOW THE INSTITUTE WAS CONDUCTED

3. On the whole, the course was conducted
   ____ Very well
   ____ Fairly well
   ____ Poorly
   ____ Very poorly

4. Lecture and discussion
   ____ Too much lecture
   ____ Too much discussion
   ____ About the right amount of each

5. Discussion leaders
   ____Too many from the University
   ____ Too many from business and industry
   ____ O.K.

6. Visual aids
   ____ Not enough movies, charts, etc.
   ____ Too much use of demonstrations, blackboards, movies, charts, etc.
   ____ O.K.

APPLICATION OF THE COURSE

7. Did the Institute apply to your particular operations?
   ____ Yes ____ Partly ____ No

FOLLOW-UP

8. Would you like to attend another institute?
   ____ Yes ____ No

9. These Institutes should run for ____ 5 days, ____ 4 days, ____ 3 days.

Fig. 4. Supervisor Institute Program Evaluation Form (Continued on page 42)

Also, conferees who enjoy a training program are more likely to obtain maximum benefit from it. According to Spencer, "for maximum learning you must have interest and enthusiasm." In a talk given by Cloyd Steinmetz of Reynolds Metals and past president of ASTD, he stressed: "It is not enough to say , 'Boys, here is the information, take it!' We must make it interesting and motivate them to want to take it."

10. Please list 3 of your main problems:

1. _____

_____

2. _____

_____

3. _____

_____

11. Comments or suggestions

_____

_____

_____

_____

_____

_____

Fig. 4. (Completed)

To evaluate effectively, training directors should begin by doing a good job of measuring reactions and feelings of people who participate. It is important to do so in an organized fashion using written comment sheets which have been designed to obtain the desired reactions. It is also strongly suggested that the form be so designed that the comments can be tabulated and quantified. In the experience of the staff of the Management Institute, it is also desirable to have the coordinator, training director, or another trained observer make his own appraisal of the session in order to supplement the reactions of enrollees. The combination of these two evaluations is more meaningful than either one by itself.

Companies who send their people to attend outside institutes and conferences should make an effort to evaluate the reactions to these programs. Several suggested forms have been described.

When a training director has effectively measured the *reactions* of conferees and finds them to be very favorable, he can feel very proud. However, he should also feel humble because his evaluation measurement has only begun. Even though he has done a masterful job of measuring the reaction of the group, he still has no assurance that any learning has taken place. Neither has he any indication that the behavior of the participants will change because of the training program. And still farther away is any realistic way of judging which results can be attributed to the training program.

STEP 2—LEARNING

It is important to recognize that favorable reaction to a program *does not assure* learning. All of us have attended meetings in which the conference leader or speaker used enthusiasm, showmanship, visual aids, and illustrations to make his presentation well accepted by the group. A careful analysis of the subject content would reveal that he said practically nothing of value—but he did it very well.

Learning Defined

There are several definitions for learning. For the purpose of this chapter, learning is defined as follows: the principles, facts, and techniques which were understood and absorbed by the conferees. In other words, it does not include the on-the-job use of these principles, facts, and techniques. This application will be discussed later in this chapter under "Behavior."

Guideposts for Evaluating in Terms of Learning

Several guideposts should be used in establishing a procedure for measuring the amount of learning that takes place.

1. The learning of *each conferee* should be measured so that quantitative results can be determined.
2. A before-and-after approach should be used so that any learning can be related to the program.
3. As far as possible, the learning should be measured on an *objective* basis.
4. Where possible, a control group (not receiving the training) should be used to compare with the experimental group which receives the training.
5. Where possible, the evaluation results should be analyzed statistically so that learning can be proved in terms of correlation or level of confidence.

These guideposts indicate that evaluation in terms of learning is much more difficult than evaluation in terms of reaction as described earlier. A knowledge of statistics, for example, is desirable. In many cases, the training department will have to call on the assistance of a statistician to help plan the evaluation procedures, analyze the data, and interpret the results.

Suggested Methods

Classroom performance. It is relatively easy to measure the learning that takes place in training programs that are teaching skills. The following programs would fall under this category: job instruction training, work simplification, interviewing skills, induction techniques, reading improvement, effective speaking, and effective writing. Classroom activities such as demonstrations, individual performance of the skill being taught, and discussions following a role-playing situation can be used as evaluation techniques. The training director can organize these in such a way that he will obtain a fairly objective evaluation of the learning that is taking place.

For example, in a course that is teaching job instruction training (JIT) to foremen, each foreman will demonstrate in front of the class the skills of JIT. From his performance, the training director can tell whether or not he has learned the principles of JIT and can use them, at least in a classroom situation. In a work simplification program, each conferee can be required to fill out a "flow-process chart," and the training director can determine whether or not he knows how to do it. In a reading improvement program, the reading speed and comprehension of each participant can be readily determined by his classroom performance. In an effective speaking program, each conferee is normally required to give a number of talks and an alert training director can evaluate the amount of learning that is taking place by observing his successive performances.

So in these kinds of situations, an evaluation of the learning can be built into the program. If it is organized and implemented properly, the training director can obtain a fairly objective measure of the amount of learning that has taken place. He can set up before-and-after situations in which each conferee demonstrates whether or not he knows the principles and techniques being taught. In every program, therefore, where skills of some kind are being taught, the training director should plan systematic classroom evaluation to measure the learning.

Paper-and-pencil tests. Where principles and facts are taught rather than techniques, it is more difficult to evaluate learning. The most common technique is the paper-and-pencil test. In some cases, standardized tests can be purchased to measure learning. In other cases, the training director must construct his own.

To measure the learning in human relations programs, two standardized tests are quite widely used in business and industry. The first is *How Supervise?* by File and Remmers. This is published by the Psychological Corporation of New York and has been used by a number of companies on a before-and-after basis to measure the learning that takes place. A newer test is the *Supervisory Inventory on Human Relations*, by Kirkpatrick and Planty. Sample test items from the latter are listed in Fig. 5 (answered by circling A for agree or DA for disagree).

There are also standardized tests available in such areas as creativity and economics. In following the guideposts that were suggested in the beginning of this article, this kind of standardized test should be used in the following manner:

1. The test should be given to all conferees prior to the program.
2. If possible, it should also be given to a control group which is comparable to the experimental group.
3. These pretests should be analyzed in terms of two approaches. In the first place, the total score of each person should be tabulated. Secondly, the responses to each item of the inventory should be tabulated in terms of right and wrong answers. This second tabulation not only enables a training man to evaluate the program but also gives him some tips on the knowledge and understanding of the group prior to the program. This means that in the classroom he can stress those items most frequently misunderstood.
4. After the program is over, the same test or its equivalent should be given to the conferees and also to the control group. A comparison of pre-test and post-test scores and responses to individual items can then be made. A statistical analysis of these data will reveal the effectiveness of the program in terms of learning.

One important word of caution must be stated. Unless the test or inventory accurately covers the material presented, it will not be a valid measure of the effectiveness of the learning. Frequently a standardized test will cover only part of the material presented in the course. Therefore, only that part of the course covered in the inventory is being evaluated. Likewise, if certain items on the inventory are not being covered, no change in these items can be expected.

There are also many examples where training directors and others responsible for programs have developed their own paper-and-pencil tests to measure

Please Answer All Statements Even If You Are Not Sure

Copyright © 1965 by D. L. Kirkpatrick

Published by Dr. D. L. Kirkpatrick, 4380 Continental Dr., Brookfield, Wis. 53005

1. Anyone is able to do almost any job if he tries hard enough.	A	DA
2. Intelligence consists of what we've learned since we were born.	A	DA
3. If a supervisor knows all about the work to be done, he is therefore qualified to teach others how to do it.	A	DA
4. A well trained working force is a result of maintaining a large training department.	A	DA
5. In making a decision, a good supervisor is concerned with his employees' feelings about the decision.	A	DA
6. The supervisor is closer to his employees than he is to management.	A	DA
7. The best way to train a new employee is to have him watch a good employee at the job.	A	DA
8. Before deciding on the solution to a problem, a list of possible solutions should be made and compared.	A	DA
9. A supervisor should be willing to listen to almost anything the employees want to tell him.	A	DA

Fig. 5. Supervisory Inventory on Human Relations

learning in their programs. For example, the American Telephone and Telegraph Company has incorporated into their Personal Factors in Management program a short test measuring trainee sensitivity and empathy. First, each individual is asked to rank, in order of importance, 10 items dealing with human relations. The participants are then assigned to groups which work 15 minutes at the task of arriving at a group ranking of the 10 statements. Following this 15-minute "heated discussion," each individual is asked to complete a short inventory which includes the following questions:

1. (a) Were you satisfied with the performance of the group?
   Yes____ No____
   (b) How many will say that they were satisfied with the performance of the group? ____
2. (a) Do you feel that the discussion was dominated by two or three members? Yes____ No____
   (b) How many will say that they thought the discussion was dominated by two or three members? ____
3. (a) Did you have any feelings about the items being ranked that, for some reason, you felt it wise not to express during the discussion?
   Yes____ No____
   (b) How many will say that they had such feelings? ____
4. (a) Did you talk as often as you wished to in the discussion?
   Yes____ No____
   (b) How many will say that they spoke as often as they wished? ____

The successive class sessions then attempted to teach each conferee to be more sensitive to the feelings and ideas of other people. Later in the course,

another "empathy" test was given to see whether there was an increase in sensitivity.

In Morris A. Savitt's article called "Is Maagement Training Worthwhile?"[2] he described a program that he evaluated. He devised a questionnaire which was given at the beginning of the program "to determine how much knowledge of management principles and practices the conferees had at the beginning." At the end of the 10-week program, the same questionnaire was administered to test the progress made during the course. This is an example of a questionnaire tailored to a specific program.

Nile Soik of the Allen-Bradley Company described an additional evaluation procedure in his article in the March, 1968, issue of the *Journal of the ASTD.* Not only did he use the *Supervisory Inventory on Human Relations* before and after the program, but he also administered it six months later. He was measuring the forgetting that took place in the period following the program.

And so we see that the paper-and-pencil test can be used effectively in measuring the learning that takes place in a training program. It should be emphasized again that the approach to this kind of evaluation should be systematic and statistically oriented. A comparison of before-and-after scores and responses can then be made to prove how much learning has taken place.

Conclusions about Learning

It is easy to see that it is much more difficult to measure *learning* than it is to measure *reaction* to a program. A great deal of work is required in planning the evaluation procedure, in analyzing the data that are obtained, and in interpreting the results. Wherever possible, it is suggested that training directors devise their own methods and techniques. As has been pointed out in this section, it is relatively easy to plan classroom demonstrations and presentations to measure learning where the program is aimed at the teaching of skills. Where principles and facts are the objectives of the training program, it is advisable to use a paper-and-pencil test. Where suitable standardized tests can be found, it is easier to use them. In many programs, however, it is not possible to find a standardized test and the trainig man must use his skill and ingenuity in devising his own measuring instrument.

If a training director can prove that his program has been effective in terms of learning as well as in terms of reaction, he has objective data to use in selling future programs and in increasing his status and position in the company.

STEP 3—BEHAVIOR

A personal experience may be the best way of introducing this section. When I joined the Management Institute of the University of Wisconsin in 1949, one of my first assignments was to sit through a one-week course called "Human Relations for Foremen and Supervisors." During the week I was particularly impressed by a foreman named Herman from a Milwaukee company. Whenever a conference leader asked a question requiring a good understanding of human relations principles and techniques, Herman was the first one who raised his hand. He had all the answers in terms of good human relations approaches. I was very much impressed and I said to myself, "If I were in industry, I would like to work for a man like Herman."

It so happened that I had a first cousin who was working for that company. And oddly enough Herman was has boss. At my first opportunity, I talked

with my cousin, Jim, and asked him about Herman. Jim told me that Herman might know all the principles and techniques of human relations, but he certainly did not practice them on the job. He performed as the typical "bull-of-the-woods" who had little consideration for the feelings and ideas of his subordinates. At this time I began to realize there may be a big difference between knowing principles and techniques and using them on the job.

Robert Katz, professor at Dartmouth, wrote an article in the July-August, 1956, issue of the *Harvard Business Review*. The article was called "*Human Relations Skills Can Be Sharpened.*" In it he stated that if a person is going to change his job behavior, five basic requirements must exist:

1. He must want to improve.
2. He must recognize his own weaknesses.
3. He must work in a permissive climate.
4. He must have some help from someone who is interested and skilled.
5. He must have an opportunity to try out the new ideas.

It seems that Katz has put his finger on the problems that exist in a transition between learning and changes in behavior on the job.

Evaluation of training programs in terms of on-the-job behavior is more difficult than the reaction and learning evaluations described in the two previous sections. A more scientific approach is needed, and many factors must be considered. During the last few years a number of attempts have been made, and more and more effort is being put in this direction.

Several guideposts are to be followed in evaluating training programs in terms of behavioral changes:

1. A *systematic* appraisal should be made of on-the-job performance on a *before-and-after* basis.
2. The appraisal of performance should be made by one or more of the following groups (the more the better):
   a. The person receiving the training.
   b. His superior or superiors.
   c. His subordinates.
   d. His peers or other people thoroughly familiar with his performance.
3. A statistical analysis should be made to compare before-and-after performance and relate changes to the training program.
4. The post-training appraisal should be made three months or more after the training so that the trainees have an opportunity to put into practice what they have learned. Subsequent appraisals may add to the validity of the study.
5. A control group (not receiving the training) should be used.

Some of the best evaluation studies are briefly described below.

The Fleishman-Harris Studies[3]

To evaluate a training program that has been conducted at the central school of the International Harvester Company, Fleishman developed a study

design and a battery of research instruments for measuring the effectiveness of the training. Seven paper-and-pencil questionnaires were used and the trainees, their superiors, and their subordinates were all surveyed.

To supplement the data that Fleishman had discovered, Harris conducted a follow-up study in the same organization. He used a before-and-after measure of job performance and worked with experimental and control groups. He obtained information from the trainees themselves as well as from their subordinates.

Survey Research Center Studies[4]

The Survey Research Center of the University of Michigan has contributed much to evaluation of training programs in terms of on-the-job behavior. To measure the effectiveness of a human relations program conducted by Dr. Norman Maier at the Detroit Edison Co., and to measure the results of an experimental program called "feedback," a scientific approach to evaluation was used. A basic design was to use a before-and-after measure of on-the-job performance with experimental as well as control groups. The supervisors receiving the training as well as their subordinates were surveyed in order to compare the results of the research. The instrument used for measuring these changes was an attitude and opinion survey designed and developed by the Survey Research Center.

The Lindholm Study[5]

This study was carried out in the home office of a small insurance company during the period of October, 1950, to May, 1951. A questionnaire developed as part of the research program of the Industrial Relations Center of the University of Minnesota was used. It was given on a before-and-after basis to the subordinates of those who took the training. No control group was used. A statistical analysis of the before-and-after results of the attitude survey determined the effectiveness of the program in terms of on-the-job behavior.

The Blocker Study[6]

A different approach was used in the study conducted in an insurance company having approximately 600 employees. Fifteen supervisors who took a course on democratic leadership were analyzed during the three-month period following the course. Eight of the supervisors were classified as authoritarian based on their behavior prior to the program.

During the three-month period immediately following the program, the changes in behavior of the supervisors were analyzed through a study of their interview records. They used standard printed forms which made provision for recording the reason for the interview attitude of the employee, comments of the supervisor, and action taken, if any. Each supervisor was required to make a complete record of each interview. They did not know that these records were to be used for an evaluation study. There were a total of 376 interviews with 186 employees.

The interview records were classified as authoritarian or democratic. The changes in interview approach and techniques were studied during the three-month period following the course to determine if on-the-job behavior of the supervisors changed.

The Tarnopol Approach[7]

In his article called "Evaluate Your Training Program," Tarnopol suggests the approach to use as well as a specific example of an evaluation experiment. He believes in the employee-attitude survey given on a before-and-after basis using control as well as experimental groups. He stresses that "in our experiences, five employees are a good minimum for measuring the behavior of their supervisor." He also stresses that "although canned questionnaires are available, it is advisable to use measuring instruments that are specifically suited to the requirements of both your company and your training program."

In his employee-attitude approach, Tarnopol has suggested inserting some neutral questions which do not relate to the training being given. This is an added factor in interpreting the results of the research.

The Moon-Hariton Study[8]

Their study was made in an engineering section of a department of the General Electric Company. The staff of the General Electric Company was assisted by a representative of The Psychological Corporation.

Two years after the adoption of a new appraisal and training program, a decision was made to attempt to evaluate its effectiveness. It was felt that the opinion of the subordinates about changes in the managers' attitudes and behavior would provide a better measure than what the managers themselves thought about the benefits of the program. Thus a questionnaire was designed to obtain the subordinates' views about changes in their managers. Nevertheless, it was felt that the opinions of the manager would add to the picture. Accordingly, they were also surveyed.

The questionnaire asked the respondents to compare present conditions with what they were two years ago. In other words, instead of measuring the attitudes before and after the program, the subordinates and the managers were asked to indicate what changes had taken place during the last two years.

The Buchanan-Brunstetter Study[9]

At the Republic Aviation Corporation, an attempt was made to measure the results of a training program. The questionnaire was used, and an experimental and a control group were measured. The experimental group had received the training program during the past year while the control group was going to receive it during the following year. The subordinates of the supervisors in each one of these groups were asked to complete a questionnaire which related to the on-the-job behavior of their supervisor. After answering the questionnaire in which they described the job behavior of their supervisor, they were asked to go over the questionnaire again and to place a check opposite any items "(1) which you think are *more* effectively done now than a year ago; (2) which you think are *less* effectively done now than a year ago."

In this experiment as well as in the Moon-Hariton approach, the subordinates were asked to indicate what changes in behavior had taken place during the last year. This was done because a measure of their pre-program behavior had not been made.

The Stroud Study[10]

A new training program called personal Factors in Management was evalu-

ated at the Bell Telephone Company of Pennsylvania by Peggy Stroud. Several different approaches were used to compare the results and obtain a more valid indication of on-the-job behavioral changes that resulted from the program. The first step was a formulation of a questionnaire to be filled out by four separate groups: (1) conferees, (2) controllees (supervisors not taking the course), (3) superiors of the conferees, and (4) superiors of the controllees.

The first part of the questionnaire was the consideration scale taken from the leader behavior description questionnaire originated in the Ohio State leadership studies. The second part of the questionnaire was called the "critical incident" section in which the conferee and control groups were asked to describe four type of incidents that had occurred on the job. The third and final section of the questionnaire applied to the conferees only. They were asked to rate the extent to which they felt the training course had helped them achieve each of its five stated objectives.

It was decided to conduct an extensive evaluation of the training program after the program had begun. Therefore it was not possible to make a before-and-after comparison. In this study, an attempt was made to get the questionnaire respondents to compare on-the-job behavior before the program with that following the program. According to Stroud, it would have been better to measure behavior prior to the program and then compare it to behavior measured after the program.

This study called "Evaluating A Human Relations Training Program" is one of the best attempts this writer has discovered. The various evaluation results are compared and fairly concrete interpretations made.

The Sorensen Study[11]

One of the most comprehensive research studies that has been done to evaluate the effectiveness of a training program in terms of on-the-job behavior was made at the Crotonville advanced management course of the General Electric Company. It was called the "Observed Changes Enquiry."

The purpose of the "enquiry" was to answer these questions:

1. Have manager graduates of General Electric's advanced management courses been observed to have changed in their manner and managing?
2. What inferences may be made from similarities and differences of changes observed in graduates and nongraduates?

First of all, the managers (graduates and nongraduates alike) were asked to indicate changes they had observed in their own manner of managing during the previous twelve months. Secondly, subordinates were asked to describe changes they had observed in the managers during the past twelve months. Thirdly, their peers (looking sideways) were asked to describe changes in behavior. And finally, the superiors of the control and experimental groups were asked to describe the same changes in behavior. This gave Sorensen an excellent opportunity to compare the observed changes in all four groups.

In this extensive research, Sorensen used experimental as well as control groups. He also used four different approaches to measure observed changes. These include the man himself, his subordinates, his peers, and his superiors. In this research, he did not use a before-and-after measure but rather asked each of the participants to indicate what changes, if any, had taken place during the past year.

Conclusions about Behavior

These are some of the best approaches that have been used to measure effectiveness of training programs in terms of on-the-job behavior. Only the methods and instruments used in these studies have been mentioned. The results, although interesting, cannot be borrowed by other training directors, but the techniques can.

For those interested in evaluating in terms of behavioral changes, it is strongly suggested that these studies be carefully analyzed. The references in this chapter indicate where the detailed articles can be found.

Once more I would like to emphasize that the future of training directors and their programs depends to a large extent on their effectiveness. To determine effectiveness, attempts must be made to measure in scientific and statistical terms. Measuring changes in behavior resulting from training programs involves a very complicated procedure. But it is worthwhile and necessary if training programs are going to increase in effectiveness and their benefits made clear to top management.

It is obvious that very few training directors have the background, skill, and times to engage in extensive evaluations. It is therefore frequently necessary to call on statisticians, research people, and consultants for advice and help.

STEP 4—RESULTS

The objectives of most training programs can be stated in terms of results such as: reduced turnover; reduced costs; improved efficiency; reduction in grievances; increase in quality and quantity of production; or improved morale which ,it is hoped, will lead to some of the previously stated results. From an evaluation standpoint, it would be best to evaluate training programs directly in terms of results desired. There are, however, so many complicating factors that it is extremely difficult, if not impossible, to evaluate certain kinds of programs in terms of results. Therefore, it is recommended that training directors evaluate in terms of reaction, learning, and behavior.

Certain kinds of training programs, though, are relatively easy to evaluate in terms of results. For example, in teaching clerical personnel to do a more effective typing job, you can measure the number of words per minute on a before-and-after basis. If you are trying to reduce grievances in your plant, you can measure the number of grievances before and after the training program. If you are attempting to reduce accidents, a before-and-after measure can be made. One word of caution, however. E. C. Keachie stated it as follows in an issue of the *Journal of Industrial Training* (July-August, 1948): "Difficulties in the evaluation of training are evident at the outset in the problem technically called 'the separation of variables'; that is, how much of the improvement is due to training as compared to other factors?" This is the problem that makes it very difficult to measure results that can be attributed directly to a specific training program.

Below are described several evaluations that have been made in terms of results. They do not offer specific formulas for other training directors to follow, but they do suggest procedures and approaches which can be effectively used.

Safety Programs

Many attempts have been made to evaluate the effectiveness of safety training programs in terms of lost-time accidents. One study was conducted by

Philip E. Beekman, plant administrator of salaried personnel for the Colgate-Palmolive Company, Jersey City plant. This study was briefly described in the Number 3, 1958, *Supervisory Management Newsletter* of the American Management Association. A comparison was made of plant safety records for the nine-month period before the training program with a comparable period after the program. The frequency rate for lost-time accidents was measured along with the number of reported accidents. The frequency rate dropped from 4.5 percent to 2.9 percent and the number of reported accidents dropped from 41 to 32. This improvement was credited directly to the training effort because no physical changes were made which affected the accident rate.

At a recent conference at the University of Wisconsin, Dr. G. Roy Fugal of the General Electric Company described a before-and-after evaluation of one of their safety programs. The purpose of the training was to reduce the number of accidents and to increase the regularity with which all accidents, major and minor, were reported. The training program consisted of the usual presentations, discussions, and movies which were very dramatic in describing accidents and their implications. The comprehensive evaluation indicated that the training program did not have desirable results. Therefore, a new approach to training was adopted which was more oriented to the job relationship between the foreman and each worker. An evaluation of this kind of safety training program did indicate the desired results.

Postal Carrier Training

In the September-October, 1957, issue of the *Journal of the ASTD*, John C. Massey described a program in which he evaluated in terms of results. Experimental group A received 35 hours of orientation training under the post office training and development program. A comparable group called control group B did not receive any training. Results of this study are shown in Table 1.

The design of this evaluation study includes an experimental as well as a control group. The importance of using these was emphasized in step 3.

Table 1. Results of Training

Category	Total Number of Incidents	
	Experimental Group A	Control Group B
Negligent accidents	5	8
Misdeliveries	21	33
Mishandling valuable mail	3	7
Late reporting	35	32
Absence without reporting	3	6
Abuse of sick leave	8	12
Errors in relay operations	13	22
Adverse probationary reports	4	9
Discourtesy	4	5

It should also be used in evaluating results wherever possible to overcome the difficulty described by Dr. Keachie.

An Insurance Company Study

In a recent letter, S. W. Schallert of the Farmers Mutual Insurance Company of Madison, Wisconsin, reported to me on an evaluation he had made. A number of their claims adjustors were enrolled in the Vale Technical Institute of Blairsville, Pennsylvania. The purpose of the three-week course was to improve the ability of adjustors to estimate and appraise automobile physical damage.

The specific technique used by Schallert was to have the adjustors keep track of their savings for approximately six months after returning from Vale. These savings were the difference between the estimate of damage by garages and the estimate of damage by the claims adjustors who had been trained at Vale. Where the final cost of the adjustment was the same as the estimate made by the Farmers Mutual man, this was considered the savings.

In other words, the purpose of the training was to prepare the adjustors to make estimates which they could justify and sell. Actual dollars and cents figures could then be used to determine whether or not the cost of sending these adjustors to Vale was justified.

A Cost Reduction Institute

Several years ago, two graduate students at the University of Wisconsin attempted to measure the results of a "Cost Reduction Institute" conducted by the Management Institute of the university. Two techniques were used. The first was to conduct depth interviews with some of the supervisors who had attended the course and with their immediate superiors. The other technique was to mail questionnaires to the remaining enrollees and to their supervisors. Following is a brief summary of that study:

DEPTH INTERVIEWS

Interview with Trainees

1. Have you been able to reduce costs in the few weeks that you have been back on the job?

Replies:	13 men	Yes
	3 men	No
	2 men	Noncommittal or evasive
	1 man	Failed to answer

2. How? What were the estimated savings?

Different types of replies indicated that the thirteen people who said they had made cost reductions had done so in different areas. But their ideas stemmed directly from the program, according to these trainees.

Interview of Superiors

Eight of the cost reduction actions described by the trainees were confirmed by the immediate superior and these superiors estimated total savings to be from $15,000 to $21,000 per year. The specific ideas that were used were described by superiors as well as by the trainees.

MAILED QUESTIONNAIRES

Questionnaires were mailed to those trainees who were not contacted personally. The results on the questionnaire were not nearly as specific and useful as the ones obtained by personal interview. The study concluded that it is probably better to use the personal interview rather than a questionnaire to measure this kind of program.

Measuring Organizational Performance

Another sophisticated and penetrating article related to evaluation was written by Rensis Likert. It appeared in the March-April, 1958, issue of the *Harvard Business Review*. It shows how changes in productivity can be measured on a before-and-after basis. Two different types of groups were used. The first was a group of supervisors trained in using a democratic kind of leadership in which decision making involved the participative technique. The supervisors in the other group were trained to make their own decisions and not ask subordinates for suggestions.

In addition to measuring the results in terms of productivity, such factors as loyalty, attitudes, interest, and work involvement were also measured. Where both training programs resulted in positive changes in productivity, the "participative" approach resulted in better feelings, attitudes, and other human relations factors.

The article described another excellent study from the University of Michigan. Dr. Likert concluded by saying that "industry needs more adequate measures of organizational performance than it is now getting."

Conclusions about Results

The evaluation of training programs in terms of "results" is progressing at a very slow rate. Where the objectives of training programs are as specific as the reduction of accidents, the reduction of grievances, and the reduction of costs, we find a number of attempts have been made. In a few of them, the researchers have attempted to segregate factors other than training which might have had an effect. In most cases, the measure on a before-and-after basis has been directly attributed to the training even though other factors might have been influential.

Studies like those of Likert attempt to penetrate the difficulties encountered in measuring such programs as human relations, decision making, and the like. In the years to come, we will see more efforts along this direction, and eventually we may be able to measure human relations training, for example, in terms of dollars and cents. At the present time, however, our research techniques are not adequate.

SUMMARY

One purpose of this chapter is to stimulate training people to take a penetrating look at evaluation. Their own future and the future of their programs depend to a large extent on their ability to evaluate and use evaluation results.

Another objective has been to clarify the meaning of evaluation. By breaking it down into reaction, learning, behavior, and results, the training person can begin to do something about it and can gradually progress from a simple subjective reaction sheet to a research design that measures tangible results.

Articles on evaluation will continue to appear in the *Training and Development Journal* and other magazines. Some of these articles are well worth studying because they describe effective principles, procedures, and methods of evaluation. One example of an excellent article is "Performance Oriented Training: Results Measurement and Follow-up," by Lerda and Cross, which appeared in the August, 1962, *Journal of the ASTD.* Another fine article appeared in the October, 1962, *Journal of the ASTD.* It was entitled "The Development and Evaluation of a Programmed Training Course for Sales Personnel," by J. S. Abma. These articles are worthwhile because they provide helpful principles and approaches which other training people can borrow. Likewise, many of the evaluation articles that continue to appear aren't worth the time it takes to read them. A list of other worthwhile reading appears at the end of this chapter under "References."

This brief chapter has not provided the answers to the training director's problem of evaluation. It has attempted to provide an understanding of principles and methods. Better understanding will come from continued study of new principles and methods that are described in articles written in professional journals. Needless to say, skill in using proper evaluation methods can only come from practice.

References

1. Goodacre, Daniel M., III: "The Experimental Evaluation of Management Training: Principles and Practice," *Personnel,* B. F. Goodrich Company, May, 1957.

2. *Personnel,* September-October, 1957, American Management Association.

3. Fleishman, E. A., E. F. Harris, and H. E. Buntt: "Leadership and Supervision in Industry," Personnel Research Board, Ohio State University, Columbus, Ohio, 1955, p. 110.

4. Mann, Floyd: "Human Relations in the Industrial Setting," Survey Research Center, University of Michigan, Ann Arbor, Mich.

5. Lindholm, T. R.: "Supervisory Training and Employee Attitudes," *Journal of the ASTD,* November-December, 1953.

6. Blocker, C. E.: "Evaluation of a Human Relations Training Course," *Journal of the ASTD,* May-June, 1955.

7. Tarnopol, Lester: "Evaluate Your Training Program," *Journal of the ASTD,* March-April, 1957.

8. Moon, C. G., and Theodore Hariton: "Evaluating an Appraisal and Feedback Training Program," *Personnel,* November-December, 1958.

9. Buchanan, P. C., and P. H. Brunstetter: "A Research Approach to Management Improvement," *Journal of the ASTD,* January-February, 1959.

10. Stroud, P. V.: "Evaluating a Human Relations Training Program," *Personnel,* November-December, 1959.

11. Sorensen, Olav: "The Observed Changes Enquiry," Manager Development Consulting Service, General Electric Company, Crotonville, N. Y., May 15, 1958.

Bibliography

"Evaluation of Effective Listening," Basic Systems Incorporated, 1 East Wacker Drive, Chicago, Ill. 60601. A booklet based on a University of Michigan program.

Besco, Tiffin, and King: "Evaluation Techniques for Management Development Programs," *Journal of the ASTD,* October, 1959.

Blake and Mouton: "International Managerial Grids," *Training Directors Journal,* May, 1965.

Goodacre, D. M.: "Experimental Evaluation of Management Training: Principles and Practice," Personnel, May, 1957.

Helms, W. C.: "Shop Supervisors Evaluate Supervisory Training," *Journal of the ASTD*, January-February, 1955.

Kirkpatrick, D. L.: "How to Start an Objective Evaluation on Your Training Program," *Journal of the ASTD*, May-June, 1956.

Malouf, L. G.: "Managerial Grid Evaluated," *Training and Development Journal*, March, 1966.

Messer, Elizabeth: "Assessing and Reporting Training Needs and Progress," Personnel Methods Series No. 3, June, 1956. Booklet available from U.S. Civil Service Commission, Washington 25, D. C.

Moffie, Calhoon, and O'Brien: "Evaluation of a Management Development Program," *Personnel Psychology*, Winter, 1964.

Mosel, J. N.: "Why Training Programs Fail to Carry Over," The University of Wisconsin Extension, Department of Commerce, 600 West Kilbourn Avenue, Milwaukee, Wisc.

Murdick, R. G.: "Measuring the Profit in Industry Training Programs," *Journal of the ASTD*, April, 1960.

Nevis, Smith, and Harper: "Behavior and Attitude Changes Related to Laboratory Training," *Training Directors Journal*, February, 1965.

Norman, J. H.: "Dollars and Cents Evaluation of a Training Program," *Journal of the ASTD*, October, 1959.

Stander, Townshend, and Swartz: "A Quantitative Evaluation of a Motivational Training Program for Blue Collar Workers," *Training Directors Journal*, November, 1964.

Underwood, W. J.: "Evaluation of Laboratory Method Training," *Training Directors Journal*, May, 1965.

Viteles, M. S.: "Evaluation of a Program of Humanistic Studies for Executives," *Personnel Psychology*, Spring, 1959.

ON-GOING PROGRAM EVALUATION[1]

Bryan Smith, University of Florida

Most Research and Training Centers are structured on traditional organization theory. This theory, which is based on a static concept of man, is often incompatible with the role expectations of the workers in these organizations and leads to functional inefficiency and ineffectiveness. Ironically, these are the major undesirable qualities that originally prompted the creation of the traditional model. A new organizational model which utilizes a modern dynamic concept of man is needed. Both governmental agencies and the private sector in our free enterprise system have utilized new managerial techniques to meet this need. These techniques include Contractual Requirements Recording Analysis and Management (*CRAM*); Implementation, Planning and Control Technique (*imPACT*); Least Cost Estimating and Scheduling System (*LESS*); and Program Evaluation and Review Technique (*PERT*). It is this latter method which is compatible with the structure and function of Rehabilitation Research and Training Centers. Because of this potential, the PERT process is recommended as a procedure to be utilized as an on-going program evaluation tool.

THE NEED FOR PERT

Traditional organization theory was built on a combination of an accounting and industrial engineering model. The responsibilities of the organization are rationally assigned to a group of boxes spread about in a structure resembling a family tree. It utilizes a balance system of authority and responsibility. Each box is vested with a certain amount of authority and simultaneously receives responsibility for productivity.

This balance is truly unique to this model. In no other part of life is authority equivalent to responsibility. Any arrangement whereby responsibilities are precisely and explicitly defined and limited cannot be regarded as efficient. It is expensive to the organization when a worker does exactly what he is told and no more.

[1] Paper presented at the seminar entitled, "Strategies for the Evaluation of Short-Term Training in Rehabilitation," February 1969, Rehabilitation Research and Training Center in Mental Retardation, University of Oregon.

Output is the responsibility of the hierarchical line. The staff is to advise, support, and provide expert assistance to this line. They only have authority within their own staff groups. Theoretically, if a balance is maintained between responsibility and authority, the staff which lacks authority has no responsibility! To equalize responsibility and authority demands that these commodities be measured—a task which defies logic.

Another characteristic of classical organizational theory is that it maximizes neatness and control. Breaking up the organizational goals into objectives and placing them into appropriate boxes gives the impression that one knows what ought to be done everywhere and that there is always an overview of the output for which a box is responsible. When something goes wrong, repercussions can be heard all the way down the organizational structure. There is special emphasis on error, its detection, and correction after it has happened. The standard organization is set up so that everyone has an assigned task. As soon as the task is accomplished, it is possible to discover the error, make appropriate adjustments, and then see that it is done. The upper levels of the hierarchy are relieved that they have made it through another day without anything going seriously wrong. These people at the higher levels always have a check to be sure that, contrary to what they were afraid of, nothing serious had gone wrong in the system. It does not provide the symmetrical check as to whether anything serious had gone right.

Another characteristic of the traditional model is that growth or development in the individual is discouraged since it will upset the system. The worker obtained a job that was appropriate for him. If he develops new skills, he will either want to reach out and do more or work less. Both have deleterious effects upon his and others' morale. The stability of workers, both in minimizing turnover and change within the individual, is desired.

These are a few of the characteristics which describe the traditional approach to organizational structure. What is the nature of man who fits into this system? What assumptions are made about the labor force from which we draw our workers into research and training centers? Within the traditional structure there are procedures and systems which imply to the worker that he is lazy, shortsighted, liable to make mistakes, selfish, and a little dishonest. Since the system does not let him see all the factors involved, his judgment, if he has any, is apt to be poor.

Likert (Haire, 1959) analyzed worker attitudes at various levels in the hierarchial order. He asked job foremen to rate the importance of certain job qualities as they viewed them for themselves, and for the workers at levels below and above them. The general foremen were asked to complete the same rating form for themselves and as they thought the foremen and the superintendent would do it. In each case, no matter what the level of the respondent, his job and his superiors demanded admirable personal qualities, such as initiative. His subordinates did not.

If a new decentralized organizational model were designed to replace the traditional hierarchial ladder, it should maximize participation and utilize worker initiative by creating organizational equality for everyone. In seeking initiative, there would be an atmosphere in which there is freedom to make mistakes, to communicate, to make new approaches, to learn, to grow, and to develop without someone anticipating failure. The growth and development of the workers are the primary goals. Through such a model a Hawthorne Effect will be injected, creating an improved organization and product produced by it.

PROGRAM IMPLEMENTATION THROUGH PERT

PERT is a proven management tool which enables an organization to simplify scheduling, allow for optimum use of staff time, simplify budget preparation, give advance warning of impending bottlenecks, and minimize administrative overload. It offers a number of insights into the organization's operations and provides the freedom for all workers to explore numerous alternatives for the accomplishment of program goals.

PERT was introduced by the United States Navy in 1958 to coordinate the complex design, manufacture, and assembly of the numerous components in the development of the operational Polaris missile. It was later used by other branches of the armed services and numerous industrial and governmental organizations. *PERT* consists of plotting the network of activities that lead to a specific outcome or objective. The Critical Path Method (*CPM*) was developed about the same time by DuPont and is a similar method of outlining, in a time sequence, the steps that must be taken to reach an objective. Both techniques are methods for rational planning and seek to coordinate a great many complex activities and efforts into a plan designed to achieve a specific outcome. They require the listing and ordering of events and activities to show what must be done, how long it will take, and the relationship of activities that must precede the accomplishment of the program. Calculations can be made based on time estimates to show the critical timing of the activities. Before the program is started, the probability of achieving the program objective in the time allowed can be calculated and, if necessary, information on each segment can be fed back into the system to alter the sequences.

PERT and *CPM* provide a visual estimate of what needs to be done, how long it will take to do it, what kind of activities must be carried out, when, where, and by whom. The main advantage is the provision of a mechanism which sets forth estimates of the time necessary to accomplish a sequence of activities, and simultaneously allows for rapid and clear communication between workers. It is possible for each person to see his and his co-workers' relationship to the success or failure of the total process.

PERT Steps

Developing specific outcomes (Step 1). The outcome must be tangible and should signify completion of an activity. There is nothing more to be done after this action takes place. It should not include any processes or previous actions. An outcome is called an event or milestone. It consumes neither time nor resources. Examples of events are: meeting held, report mailed, data processed, cards punched. They are occurrences which may happen during the course of the program and are not subject to evaluation.

Determining precedent events (Step 2). All the events which precede the final event must be recorded. This listing does not have to be in a serial or chronological order. For example, if the final event is: center opened, the precedent events might be: staff hired, materials ordered, equipment installed, etc.

Order precedent events (Step 3). After the events are listed, they should be ordered. To the left of each event write a number indicating the order in which these events *must* occur. For example, before hiring the Center's staff, it is necessary for the workers to be accepted by the Board of Directors. New events may have to be added to the list.

Table 1.

Ordering Preceding Events

Sequence Number	Event	Preceding Event
1	Organizational meeting held	0
2	Motel contacted for seminar space	1
3	List of lecturers completed	1
4	Honorarium approved by administration	5
5	Eight lecturers recommended	3
6	Syllabus for each seminar completed	4
7	Syllabi submitted for approval	6
8	List of trainees obtained	1
9	Coordination meeting held for lectures	3,10,11
10	Trainees enrolled	8
11	Lecturers assigned to rooms	2,4
12	Seminar started	9

Once the events have been listed in sequential order and the preceding events have been determined, they can be plotted into a flow chart. This chart consists of a system of paths of sequential events contributing to the final program objective (for description of events see Table 1). The paths are parallel in the time sequence and each represents a sequence of time. The length of the path, which represents the activities between two events, are not drawn to scale.

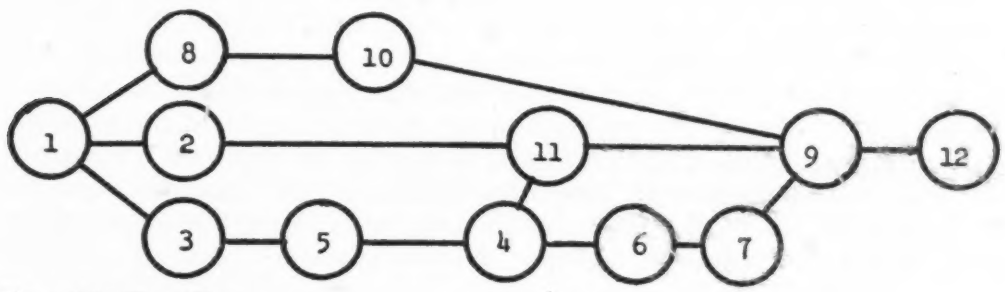

Fig. 1. The Flow Chart

Each event must be followed by a successor event. Each successor event must have at least one predecessor event. There should not be any loops, crosses, or dangles such as those illustrated in Figure 2.

Determination of activity time (Step 4). Certain activities must take place between each pair of events. These activities must be listed, and the amount of time that each would take must be estimated. In making the time estimate (t_e) three predictions are used: the minimum time (a) is the minimum possible period of time in which the activity can be accomplished. The best time (m) is the best estimate of the period of time in which the activity can be accomplished. The maximum time (b) is the maximum possible period of time it would take to accomplish the activity. These activity times are then placed into the following formula to calculate the time estimate: $t_e = \frac{a + 4m + b.}{6}$

Allocation of priority activities (Step 5). The activity time estimate is added to the flow chart on the appropriate line between the two events. Once the time estimates are on the flow charts, it is possible to determine which activities are priority activities, and which line on the chart represents the

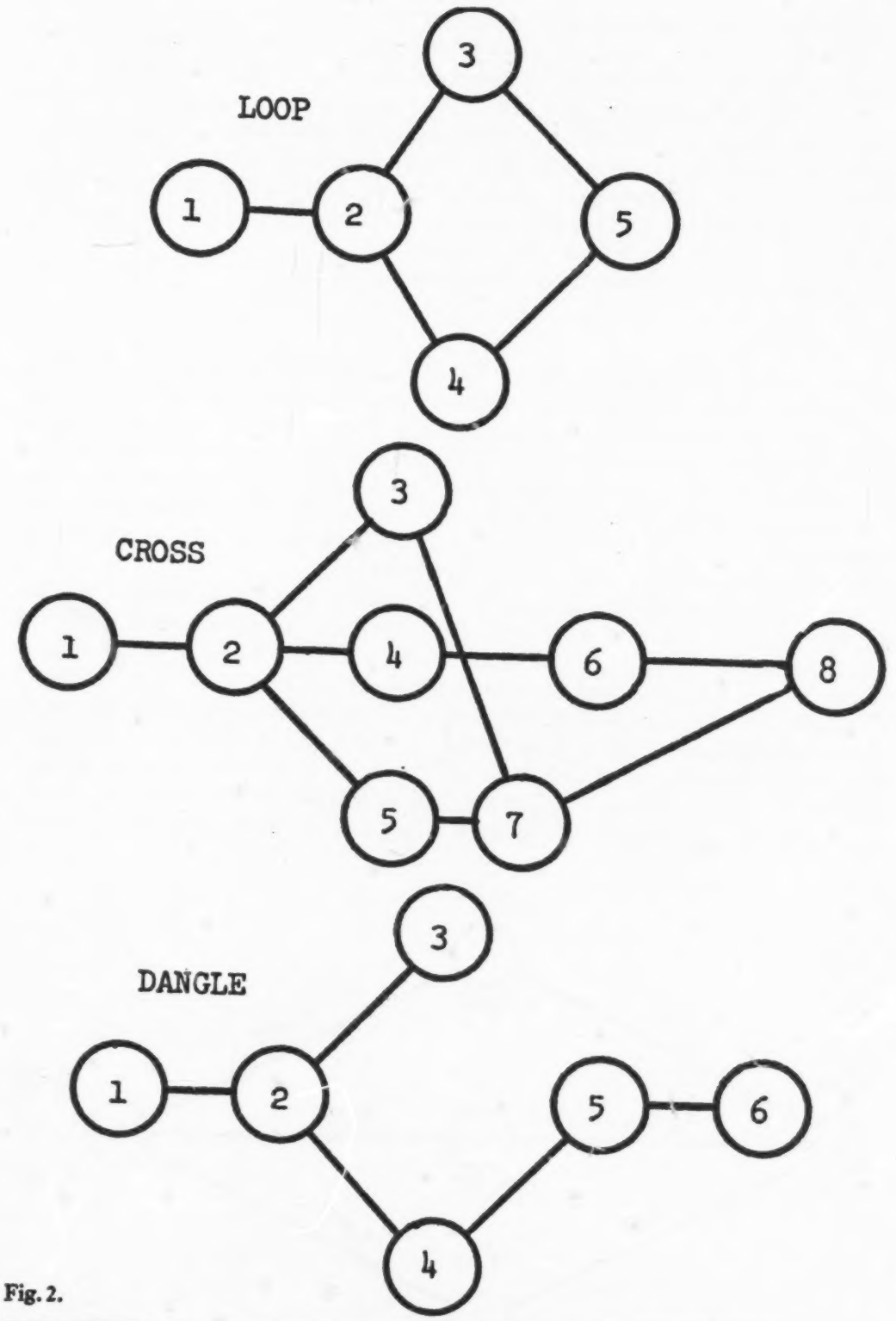

Fig. 2.

"critical path." The critical path is the longest time from the first event to the final event, and therefore, the quickest possible time the program can be completed. These are a series of events which have other events dependent on them; thus, they need priority attention. Resources and personnel can be allocated on the basis of the flow chart.

The flow chart will reveal how concurrent activities can exist and how more than one may be taking place simultaneously. These parallels can be conducted at the same time by one or more persons in the organization.

The time estimates may vary markedly between minimum and maximum. The wider the separation between these estimates, the greater the uncertainty associated with the activity. This uncertainty can be expressed by the statistical tool known as variance (σ^2). It is descriptive of the uncertainty associated with the distribution. If the uncertainty is small, the variance will be small. It shows how precise the activity time estimate is.

Variance can be calculated by the following formula:

$$\sigma^2 = \left(\frac{b-a}{2}\right)^2$$

The expected completion time (T_E) should also be calculated and added to the flow chart. It is the earliest possible time that the event can be reached and is calculated by adding the t_e of the activities leading to each event. The latest allowable completion time (T_L) can now be calculated from the T_E by subtracting the t_e. The values of all the T_L are calculated for each event and included on the flow chart below each event. In this computation start with the last event and work back toward the first one by subtracting the t_e from the value of the T_E for the successor event.

From the values of the T_E and T_L the amount of slack time (T_s) can be calculated. Slack time is the amount of time in excess of the t_e that is available to complete the activity. The slack of an event is $T_L - T_E$. Therefore, if you are given 30 days to complete an activity and can actually do it in 25 days, you have a 5-day grace period. Slack time is important in analyzing the completed project when time of completion is a vital factor. The flow chart can now reveal those areas which have an excess of manpower or material resources. It also enables the reader to spot potential trouble areas, i.e., those areas with zero or negative slack time. The value of slack, associated with an

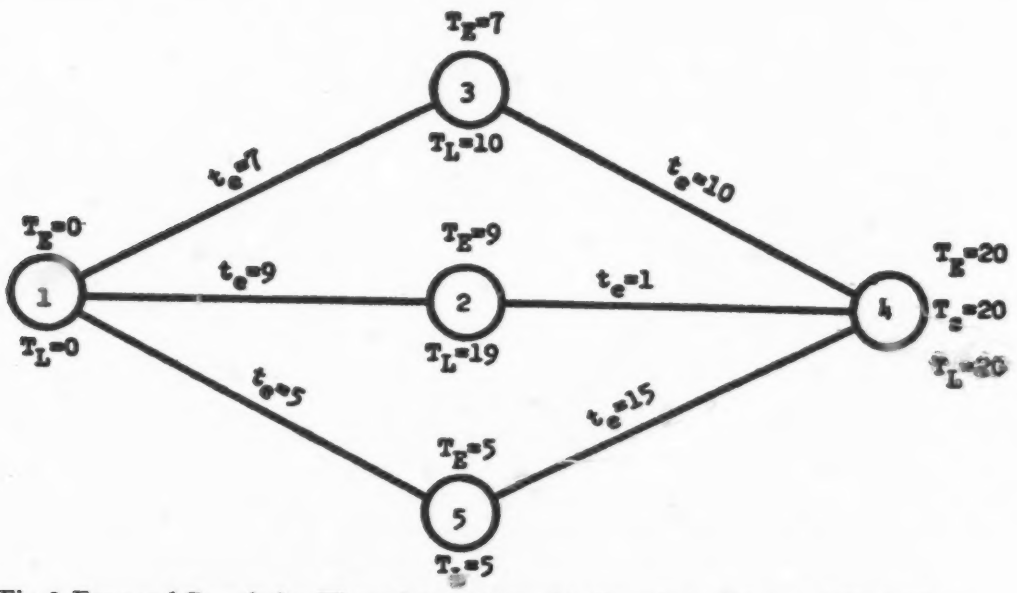

Fig. 3. Expected Completion Times, Latest Allowable Completion Time and Slack Time

event, will determine how critical that event is to the completion of the program. The smaller the slack value the more critical an event becomes.

To estimate the probability of reaching the final event in the time allowed, the following formula is used:

$$Z = \frac{T_s - T_E}{\sqrt{E\,\sigma^2 T_E}}$$

In most program flow charts there are many possible paths from the initial event to the terminal one. They will vary in the total amount of slack time. The path described as the critical path is the one with the least amount of slack time, requiring the most time to get from initial to terminal event, and causing the final event to be delayed when any event on the path slips in time.

The critical path in Figure 3 is indicated by a line from events 1-5-4.

CONCLUDING COMMENTS

The main advantage of a system such as PERT is that it forces the organization into the development of a model which shows the activities with estimates of the time necessary to accomplish the sequence of interdependent actions. The result is that priorities are established, staff time is allocated rationally, budgets are estimated more precisely, bottlenecks are eliminated, and program objectives are achieved more efficiently. The model also provides for more rapid and accurate communication between workers, divisions or agencies, since it shows the precise role of each unit in the total process. Unnecessary work is eliminated and critical work is more likely to be accomplished.

PERT is not a panacea. But, if it is applied to the decision-making process, it reveals the necessary steps about which no argument is possible, and minimizes friction in determining what activities must be conducted.

References

Avots, I. The management side of PERT. *California Management Review*, Winter 1962.

Boulanger, D. G. Program evaluation and review technique. *Advanced Management*, June-August 1961.

Clark, C. E. The optimum allocation of resources among the activities of a network. *Journal of Industrial Engineering*, January-February 1961.

Consideration of PERT assumptions. *Bulletin of the Operations Research Society of America*, Spring 1962.

Dogrusoz, H. Development programming. *Operations Research*, September-October 1961.

Federal Electric Corporation. *A programmed introduction to PERT*. New York: John Wiley and Sons, 1963.

Haire, M. *Modern organization theory*. New York: John Wiley and Sons, 1959.

Interpretations of PERT. *Harvard Business Review*, 1964, *42* (2), 160-172

Levy, F. K., Thompson, G. L., & Wiest, J. D. The abc's of the critical path method. *Harvard Business Review*, 1963 *41* (5), 98-108.

Miller, R. W. How to plan and control with PERT. *Harvard Business Review*, 1962, *40* (6), 93-104.

Paige, H. How to plan and control with PERT. *Harvard Business Review*, 1963, *41* (6), 87-95.

Problem solving through PERT. Unpublished guide by Arnold, Erickson, Phillips and Bemis, University of California at Berkeley, July 1964.

Roman, D. D. The PERT system: An appraisal of program evaluation and review technique. *Journal of Academy of Management*, April 1962.

Stilian, G. N. et al. *PERT*. New York: American Management Association, Management Series No. 14, 1962.

Stires, D. M., & Murphy, M. M. *Modern management methods of PERT and CPM*. Boston, Massachusetts: Materials Management Institute, 1962.

Teaching PERT project network techniques. *Training Directors*, December 1961.

Behavioral Criteria for Short-Term Training[1]

Leo A. Hamerlynck, University of Calgary

In part this paper is a criticism of many evaluation studies, including some of my own, but the essential point is a proposal for the solution of a major aspect of the key question in evaluation: Criterion Selection. My data for this proposal derives from diverse sources. First, I have sampled a large portion of the literature describing the evaluation of short-term training. Next has been my research activity as former training director at the University of Oregon Research and Training Center. The data from these sources has been inconsistent with my position as a psychologist of decided behavioristic views. An aspect of my behaviorism which has been particularly active in developing the inconsistency is the demand for behavioral operational definitions. The proposed system for selection of training effectiveness criteria is directly tied to the process of operational definitions in a behavioristic paradigm. I suggest that if you hold views in conflict with the behavioristic you might still consider the basic appeal to the funding public where the pragmatic, utilitarian questioner is very evident.

Rather than lead you through my experiences with the literature as well as my own research activities, I will fit these into the discussion of my criterion evaluation system. The system follows a categorization based upon: 1. *Proximity to the client population* which divides into *Primary* and *Secondary*, and 2. The *method of observation used in the collection of data*, which I divide into *Direct* and *Indirect*. The resulting system is a two-by-two paradigm.

Direct/Primary	Indirect/Primary
Direct/Secondary	Indirect/Secondary

There is a hierarchy in the system which proceeds from the premise that short-term training is conducted with a final goal, "helping people that hurt." Whether the hurt is in terms of phantom limbs or social/occupational failure,

[1] Adapted from a paper presented at the 75 Annual Convention of the American Psychological Association, September 1967, Washington, D. C., and printed with the permission of the author.

it is easy to say, "These people hurt and training programs are intended to train staff to alleviate these pains." This premise therefore dictates that the ultimate criterion of our success in training can only be found in the parameter of Primary Data—actual client behavior. The kind of observations, either Direct or Indirect, are of obvious importance, but the resultant data is judged as to significance only in relation to its proximity to client behaviors.

Direct Observation-Primary Behavior

I could not find literature under the heading of "evaluation" which described activities where the client population was directly observed. There are such activities, but not labeled as evaluative. The patients/clients of individuals trained in the application of behavior modification techniques are perfect examples of Direct/Primary category evaluation studies. The essence of such a criterion is that a person given short-term training in behavior modification techniques changes the behavior of his client. The behavior of the client served by the trainee is the criterion.

I would exclude self-report data from this category. It should be considered as primary data, but of an indirect source unless self-report behaviors are the client problem. Quite obviously, many training programs are not directed at client behavior, but at professional activities and public relations. The system will account for evaluation of such programs, but makes explicit a criterion for selection of training programs. My conservative view is that too often we lose sight of the final purpose for short-term training—client/patient welfare.

Indirect Observation-Primary Behavior

The criterion derived from this combination is also relatively rare in publications and practice. Follow-up studies using retrievable data such as releases, remission rates, employment record, etc. are used to typify this approach to evaluation. There is danger while such data may be quite adequate for use with an evaluation of a total program or system such evaluative criteria are too molar for short-term training. When self-report such as client attitudinal assessments, Q sorts, and check lists procedures are utilized, the question of behavioral change is too often left untested. The aforementioned use of client self-report or language change might be more feasible. The molarity of the former combined with the difficulty of partialing out the contribution of skills or attitudes acquired in short-term training from the regular training and treatment program leaves the Indirect/Primary classification essentially neglected.

A technique which I have found very promising is client self-recording of behavior under a behavioral contract. We used a self-recording technique with children in a special education class under contract with their teacher (Hamerlynck & Donley, 1968). Although the project resulted from a graduate seminar, the period of training in class certainly did not exceed the number of hours we use in our short-term training. The resultant feedback on my teaching effectiveness was the kind that too many instructors have never encountered: 1. I taught a professional some ways to help her students; 2. she used the techniques; and 3. problem behaviors were eliminated. The power of such evidence is self-evident and the data is hard.

The prerequisites are close observation of the target population and, of course, the content of the instruction. Expending such efforts is quickly justified by the feedback obtained.

Direct Observation-Secondary Behavior

An excellent example of evaluation activities of this description is in training using simulation activities. The observations are made directly with the trainee working on a simulation problem. The training and evaluation are essentially inseparable as the criterion measure is a vital part of the simulation.

There is a considerable gap between the behavior of a professional in a simulation problem and his client's behavior. However, for quick feedback and teaching effectiveness, the simulation approach is excellent. The reality of high cost in time and funds, if using actual client behaviors, adds to the value of simulation.

An approach of equal promise is to directly observe the trainee in his professional activity. Barclay (1968) has evaluated the effect of an NDEA Institute for School Psychologists with just such an approach. The management problems are considerable, but the information is of high significance for planning of subsequent training institutes.

Indirect Observation-Secondary Behavior

This classification dominates the evaluative activities evident in the literature. I am currently engaged in just such evaluative research of our training activities. Such data is of a secondary source, probably a questionnaire, attitude scale, content examination, or possibly several assessments all of which are obtained on the trainee. The data then is obtained indirectly, probably self-report, and about the professional in training.

Of obvious magnitude is the unexplained variance between an attitude scale for a professional and the behavior of his client. The same holds true for the gap between information sampled in a content test and the ultimate validation with client behavior. I have noted in my own research using a semantic differential technique for attitude measurement that our data hold real promise as a contribution to basic studies in attitudes toward exceptionality or deviance. But, it does not give us the evaluative information we need to modify or accelerate our program. The best evaluative data we get is in the soft form of phone calls, letters, and foremost, visits to our trainees. The visits allow us to observe client behavior during programs using techniques we have taught to the professionals. *This is Primary Data.* The other feedbacks might be indirect and subjective, but based upon client behaviors which is our terminal criterion.

The best that attitudinal measures can ever offer to the training agency is a predictive statement about client behaviors which would mean a saving in time of observation. The methodological problems in the measurement process alone are enough to negate this ideal point. If you contemplate the problems involved in obtaining validation data for trainee behavior and finally client behavior, the problems are of such a magnitude as to preclude such activities for evaluative purposes. It should remain an academic problem which social psychologists might attend to in pursuit of theoretical questions.

The same criticism is true of all self-report data from the trainees. To obtain data for the evaluation of the instructional environment, why not simply count the number of complaints or "glitter statements" and perhaps use an opinion questionnaire? But keep in mind that this is all data of very doubtful validity to your training goal and which can also be suspect in regard to reliability.

SUMMARY

The evaluation of short-term training must recognize client behavior as a final criterion. Any evaluation based upon data from observations of the individual in training should be directed at specific trainee behaviors derived from the curriculum. Simulation training holds real promise for this purpose as well as for the actual training. Finally, use generalized evaluation with all of the cautions demanded by this criterion system. Don't confuse data which is essentially of research nature with the kind of data you must have to evaluate your training program.

References

Barclay, J. R. *Changing the behavior of school psychologists: A training and rationale method.* Final Report, 1967, NDEA on Advanced Counseling for School Psychologists, California State College at Hayward, October 1968.

Hamerlynck, L. A., & Donley, M. Modification of high base rate verbal behavior by behavioral contracts. Paper read at 1968 Conference of Rocky Mountain Psychological Association, Denver, Colorado, May 1968.

Professional Opinions Regarding Curriculum Content in Short-Term Training Programs in Mental Retardation: An Evaluation Survey[1]

Patrick J. Flanigan, University of Wisconsin, Madison

The present programming trend in the provision of educational and habilitative services for mentally retarded individuals and their families has necessitated a nation-wide endeavor to create and implement a variety of both short and long-term training programs for professional and para-professional personnel working in the field of mental retardation. Since the effectiveness of the increased services and programs is, in large part, dependent upon trained professional individuals with basic competencies in the area of mental retardation, the nature of the curriculum of these training activities is of critical importance.

In reviewing the literature on the training of professional personnel in mental retardation, it becomes evident that little, if any, mention is made of 'specific' curricula preparatory areas which are deemed essential, to have had either didactic or practicum experience in, for working with the mentally retarded. Parallel to this obvious literature deficiency is the void in other sources concerning habilitation and special education training programs as they relate to specific content involved in program practices. Warren (1957), Hylbert (1963), Olshansky (1957), Cantrell (1958), Fleischer (1964), and the Kent State Study (1965) all discuss curricula for rehabilitation counselors but nowhere in their descriptive statements do they contain specific information pertaining to the mentally retarded. Each of these studies concluded that the labor market and employment relations were critically important content areas and most of them mentioned that all training courses must include practical applications and field work experiences.

Goldman (1959), Stevens (1955), Lerner (1957), Wolinsky (1967, 1958,

[1] Paper presented at the seminar entitled, "Strategies for the Evaluation of Short-Term Training in Rehabilitation," February 1969, Rehabilitation Research and Training Center in Mental Retardation, University of Oregon.

1959), Cain (1964), Olson and Hahn (1964), Fils and Attwell (1966), Blessing (1966), and Milazzo and Blessing (1964) examined programs for educators of the mentally retarded but made no references to rehabilitation practices and principles in their discourse. Stevens (1963), Harrison (1963), and Kokaska (1966) attempted to identify problems and explain programs for institutional staff and camp volunteers working with the retarded. But their information did not contain any principles relating to rehabilitation practices.

Two suggestions for future training programs recur throughout the aforementioned articles. These are: 1) more instruction should be given in the behavioral sciences, such as sociology, psychology, economics, medicine and education; and 2) additional involvement in concrete experiences is deemed absolutely necessary.

Six authors dealt, to varying extents, with the problem of bringing rehabilitation personnel and personnel working with the retarded together in a training situation. But only one of them, Clark (1967), included in her description all of the professions theoretically working with the retarded although she did not intervene during the active training processes and attempt to discover the trainees' perception of curricula needs. Younie (1966) and Meyers (1962) discussed a vocational rehabilitation curriculum for special education teachers. The teachers in Younie's study wanted more information about rehabilitation philosophy, terminology and patterns of local services as well as classroom programs and vocational guidance and counseling. The trainees at California State College, as reported by Clark (1967), represented several professional disciplines working in the rehabilitation realm along with the retarded. Their primary focus was reviewing patterns of local services and institutional visitations. Power (1962), Jaslow (1967) and Seidenfeld (1962) in their manuscripts have advocated vocational rehabilitation programs for working with the retarded, however, little information is available in their sources with regard to the content, structure, and implementation of the recommended programs.

Only Cantrell (1958), Fleischer (1964), the Kent State Project (1965), Goldman (1959), Fils and Attwell (1966) and Younie (1966) involved trainees, to a limited extent, in an attempt to substantiate their specific curricula needs. Yet, only a very small number of samplings of the trainees over a very short period of time were undertaken in this regard. Lerner (1957), Cain (1964), Warren (1957), Whitten (1957), and the two rehabilitation department publications (1966, 1967) all pointed out the critical importance of doing this. Those who did not consult active trainees were primarily interested in vocational rehabilitation counselor training programs generically and not necessarily programs concerned primarily with curricula having relevance for the mentally retarded.

Thus, this project was undertaken in recognition of the paucity of existing descriptive literature pertaining to suggested program curricula as obtained from practicing professionals involved in programs and services for the retarded.

In order for any on-going educational program to effectively serve the population that it is designed to serve, it is necessary to periodically *re-evaluate* the needs of curricula content and attempt to enrich programming in current on-going activities in an attempt to meet these needs. In addition, it is believed that solutions for enrichment and modification of existing programs may be realistically found only by consulting with those professional individuals who are most cognizant of, and continually involved in, critical curricula areas.

Consequently, the purpose of this project was to directly sample the involved personnel to determine these curriculum needs and to integrate the findings into on-going short-term training programs.

THE SURVEY

Setting

Since 1964 the Rehabilitation Research and Training Center in Mental Retardation, University of Wisconsin, Madison, has had as one of its major programs the provision of short-term training courses for professional personnel working in the area of mental retardation. The intent of these programs has been to provide additional preparation in mental retardation required by professional personnel in any of the (behavioral) rehabilitative professions presently in service and to provide an orientation to, or additional preparation in, specific areas of mental retardation.

Subjects

All trainees in attendance at short-term training sessions conducted by the University of Wisconsin Research and Training Center in Mental Retardation between September 1, 1964, and October 31, 1968, voluntarily participated in the survey. More specifically, 3299 practicing professional individuals representing 12 professional orientations were screened in regard to their specific curricula suggestions which they deemed applicable to short-term training programs in mental retardation. All of the trainees at the time of the survey initiation were engaged in working with mentally retarded clients and their families on either a full or part-time caseload basis. Table 1 presents descriptive data on the number of trainees and the professional disciplines represented by all trainees included in this survey investigation.

Table 1
Summary of Professional Disciplines
Represented at Training Institutes
September 1, 1964 to October 31, 1968

VRA Counselor	783
VRA Supervisor	173
VRA Administrator	164
Education	522
Psychology	258
Social Work	340
O.T., P.T., R.T., I.T.	435
General Administrator	197
Physicians	47
Nursing	211
Clergy	5
Others	164
Total	3299

Procedure

At the opening registration of the short-term training seminars, each trainee was requested to list ten curricula areas that they felt would be of importance

to them in their professional endeavors for the following two reasons: 1) the relevance of these curricula areas to situations which repeatedly arose in their provision of professional services to retarded clients and their families; and 2) other behavioral characteristics of the retarded which they felt additional knowledge of would be beneficial to them in their generic professional activities. Approximately 30 minutes time was alloted for collection of the obtained stimuli descriptive data.

The obtained survey data which was continuously collected at the beginning of all one, two and four week short-term training sessions was arbitrarily categorized into 15 generic content areas. Table 2 presents the generic categorized areas which were used in tabulating the obtained descriptive data.

Table 2
General Categorizing Areas Used in
Classifying Obtained Curricula Data

I.	Etiology and General Characteristics
II.	Evaluation
III.	Behavioral Techniques
IV.	Counseling
V.	Care
VI.	Recreation
VII.	Rural
VIII.	Programming and Curriculum
IX.	Vocational Rehabilitation
X.	Sheltered Workshops
XI.	Administration
XII.	Public Relations
XIII.	Research
XIV.	Therapy
XV.	Sociology and Psychology

Each general category was further subdivided through the addition of suggested stimuli data on the part of the trainees. Before any of the suggested curricula data met the criterion for inclusion in the final data tabulations it had to have been suggested by at least ten different trainees irrespective of program participation.

Additionally, a further categorization of the data was a breakdown of curricula suggested by the 12 professional disciplines represented at the training sessions. The criteria utilized in this regard (further data breakdown) was again the recording of a curricula suggestion at least ten times, plus in the further breakdown analysis suggestion by members of the same professional discipline at least ten times.

Analysis and Discussion

The arbitrary classification system was developed and utilized after the data collection was undertaken for approximately one year, and observation of the data after a year gave indications as to the satisfactory method of classifying all of the obtained and anticipated data. Additional topical subheadings were added to the configuration as additional criterion content was obtained.

A secondary categorization of the data broke down the suggested stimuli items by the professional discipline of the individuals suggesting the area of content. The criterion of an item in this breakdown, being suggested at least

ten times by members of the same professional discipline, was imposed prior to the classification of a curricula item.

The tabulated data presented in Tables 3 through 17 (see pp. 78-86) summarize the 15 categorized curricula areas in the project. For clarity purposes they will be interpreted from two points of view: 1) curricula areas relating to basic and advanced academic information in mental retardation, and 2) practical techniques that could easily be transferred to specific programming practices.

Inspection of the majority of the tables (3-17) indicates that a majority of the responses obtained from this sample could be arbitrarily classified as falling primarily in the "basic information in mental retardation" category. This interpretation is reflected in many of the categorized titles which can be readily ascribed to an orientation course outline in mental retardation. A further review of many of the subtopics under many of the headings, i.e., Table 2, topic 1; Table 3, topic 8; etc., reflect numerous content areas which are at least given some attention in the initial exposure in this respect. Additional patterning of the suggested content areas and subpoints, i.e., evaluation, counseling, etc., could provide a basis for planning other training endeavors relating to basic information in mental retardation in a variety of training settings.

As a means of discussing and interpreting the obtained data in this project for clarity purposes, the following chart is presented. This chart presents an interpretive categorization of the stimuli data in this project, categorized by level and sequenced into a proposed sequential experimental short-term training program in mental retardation.

Level I, Unit A illustrates the beginning of the proposed sequence which a trainee, again from any of the rehabilitation related professions, would experience in order to begin preparation for enriching their endeavors with the retarded. A basic orientation course in mental retardation would serve to provide a thorough introduction to mental retardation. Trainees would pursue a comprehensive study of the mentally retarded as a biological and sociocultural phenomenon. Initial study of the definition, prevalence and etiology would be followed by an intensive evaluation of physical and psychological characteristics and associated medical, social, legal, and educational-vocational considerations. Various curricula resources would be available in this regard to permit an interested trainee to complete this requirement in a number of different ways, i.e., correspondence course, programmed manual course in mental retardation, etc., thus, a trainee would not necessarily have to receive this training away from his home setting.

Level II, Unit A would provide a basic enriching aspect through a sequence of visitations to specialized programs and facilities that serve the mentally retarded. These visitations, reflected periodically in the obtained stimuli data of this project, would include traditional programming practices and services focused around the specialized area of interest of the particular trainee.

Level III, Unit A would provide for an intern training situation with specific professional specialists in various programs and services dealing with mentally retarded clients. Thus, a trainee would select a professional specialist who most closely approximates his particular interest or responsibilities in service areas in mental retardation. These individuals could be selected irrespective of program setting with the primary variable in the selection process being the specific type of programming that they are involved in and the trainee could spend a blocked period of time working with this person.

Proposed Experimental Short-Term Training Program in Mental Retardation

	Level I	Level II	Level III	Level IV
A Basic Exposure Time exposure	Orientation to Mental Retardation–Basic Exposure Correspondence Course – Programmed Manual (2 weeks)	Sequence of Visitations to Programs and Facilities for the Mentally Retarded (1 week)	Intern Training with Professional Specialists in Mental Retardation (1 week)	Active Professional Participation with Retarded Clients in Naturel Job Settings (Variable)
		SEMINAR EMPHASIS		
B Correlated Integrative Seminar Time Exposure	Integration of Experiences with the Retarded and Rehabilitation Programming and Principles (1 week)	Integration of Behavioral Observations with the Retarded (Oriented toward Learning Potential, Remediation Program Prescription – Clinical Focus) (1 week)	Composition and Creation of Rehabilitation Program Prescriptions Based on Learning and Other Behavioral Principles with the Retarded (1 week)	Consultation, Self-Evaluation and Independent Study In-Service Training Responsibilities–Home Setting (1 week)
C Additional Specialized Academic Exposure with the Retarded Time exposure	Clinical Exposure in Diagnosis and Program Prescriptions with the Retarded (Variable)	Other Specialized Coursework Relating to the Mentally Retarded, i.e., Research Design and Methodology, Personality, Behavior Modification, Etc. (Variable)	Additional Independent Study and Research with the Retarded (Variable)	Advanced Graduate Degree Work (Variable)

Level IV, Unit A reflects the active professional participation that the trainee will engage in upon his return to his permanent job situation.

Level I, Unit B would further academically orient the trainee from the frame of reference of learning and the observation of learning problems as they specifically relate to the behavioral characteristics of the mentally retarded. This additional exposure would serve the following primary function, that being, of providing advanced instruction in the general learning psychology of mental retardation. Curricula areas to be emphasized would include those of intellectual, motivational, learning, and affective-personality aspects of the retarded.

Level II, Unit B would function along the line of providing the trainee with an opportunity to pursue concentrated study in the clinical aspects of mental retardation. The trainee would concentrate primarily on the techniques and problems in behavioral evaluation and diagnostic remediation treatment problems of mentally retarded children and adults. An in depth exposure in procedures for the assessment of behavioral factors would be studied. Significant consideration would be given to the various psychological techniques dealing with the numerous behavioral problems of the retarded.

Level III, Unit B would involve the trainees in a related practicum activity (related to the content or seminar emphasis in Level II, Unit B) which would give primary emphasis to the current multidisciplinary approaches to meeting the needs of the mentally retarded individual and his family.

Level IV, Unit B would provide the individual the opportunity to relate the learnings of Levels I, II, and III, Unit B to the specific job situation in which the trainee ultimately will be returning.

Level I, Unit C would serve to provide a variety of functions. Initially, the program could include such general topics as the role of federal, state and local agencies in the area of mental retardation, the function and structure of professional organizations concerned with mental retardation. Additionally, aspects of comprehensive service programs, the role of the National Association for Retarded Children and its counterpart at the state and local levels in meeting the needs of the mentally retarded would be investigated. Of primary concern during the conduct of this type of clinical experience would be the issues that are evident in numerous curricula areas such as definition, prevalence, etiology, prevention, treatment, and the role of the mentally retarded in our societal structure. Discussion of these issues in seminar fashion would permit more detailed consideration than would be possible in the beginning levels. The seminar would include the discussion, interpretation, and program prescriptions undertaken by the trainees during the initial sequencing of information, and would include direct participation by specialists from a variety of mental retardation facilities. This approach would add an integration effect to the organization, function and contribution of mental retardation facilities. Additionally, exposure to a seminar of this type would present an opportunity to enrich the trainees' research interests and skills. Research needs would be discussed and the trainee would be provided with an experience in developing research projects and, if feasible, the discussing of experiences gained during active participation in research projects relating to mental retardation. The seminar would provide trainees an opportunity to have proposed research and/or program demonstration projects evaluated.

The proposed experimental short-term training program presents many possibilities which currently are not being focused upon in most existing programs. Realistically, there are a small number of prospective trainees who would be able to obtain sufficient release time from their normal job responsi-

bilities to take advantage of the total proposed programming sequence. The majority of the participating trainees would, at best, complete Levels I, II, III, and IV, Unit A and possibly a limited exposure to portions of Unit B. Upon completion of Units A and B including all levels, it is assumed that an individual would have developed enough knowledge and skills in the practical aspects of mental retardation to conduct in-service training programs in mental retardation in their respective job settings, thereby guaranteeing an additional training effect. The proposed scheme would also provide a basic frame of reference for the completion of sequential curricula materials, i.e., instructional monographs, audio-visual aid materials, program manuals, results of research projects; which would have direct relevance to various sections of the training program. The following list represents endeavors in this regard by the Research and Training Center at Wisconsin: 1) correspondence course—Introduction to Mental Retardation (Level I, Unit) (active); 2) An Orientation to Mental Retardation: A Programmed Manual (Level I, Unit A) (published); 3) monograph—Epidemiology of "Cultural-Familial" Mental Retardation (Level I, Unit A) (published); 4) monograph—Comprehensive Family Rehabilitation of the Mentally Retarded (Level III, Unit B) (in preparation); 5) monograph—Research on the Use of Behavior Modification in Influencing the Vocational Behavior of the Mentally Retarded Client (Level II, Unit B) (published); 6) monograph—Handbook on Behavioral Modification Techniques for Use by Rehabilitation Personnel (Level I, Unit B) (in preparation); 7) monograph—Changing Concepts in Mental Retardation (Level I, Unit A) (published); 8) monograph—Pre-Vocational Preparation for the Mentally Retarded (Level II, Unit B) (published); 9) monograph—Psychological Instruments in Use with the Mentally Retarded (Level III, Unit B) (in preparation.

SUMMARY

Innumerable possibilities exist in a number of different directions for enriching content information. Certainly, various research endeavors can be easily sequenced into appropriate programming levels; this is necessary because the basic function of short-term training programs for professional personnel have as their unified objective the coupling between research endeavors and their practical application to habilitation through interpretive mechanisms in all of our on-going programs, i.e., teaching, clinical work, and research.

Data derived from various professional disciplines are depicted in Table 18 (see pp. 87-88). This analysis is presented for review and possible utilization in the planning of specific training programs in various settings.

References

Blatt, B. The preparation of special education personnel. *Review of Educational Research,* 1966, 36 (1), 151-161.

Blessing, K. R. A state department-university approach in providing internship experiences for prospective leadership personnel in special education. *Exceptional Children,* 1966, 32 (5), 305-309.

Cain, L. F. Special education moves ahead: A comment on the education of teachers. *Exceptional Children,* 1964, 20, 211-217.

Cantrell, D. Training the rehabilitation counselor. *Personnel and Guidance Journal,* 1958, 36, 382-387.

Clark, M. F. Community teams ponder retardation. *Rehabilitation Record,* July-August, 1967, 5-6.

Cohen, J. S. New university council coordinates rehabilitation services. *Journal of Rehabilitation,* 1965, 31 (3), 23.

DiMichael, S. G., Fraenkel, W. A., Power, R. F., Jervis, G. A., Seidenfeld, M. A., Mayo, L. W., Dybwad, G., & Sloan, W. Rehabilitation of the mentally retarded. *Journal of Rehabilitation*, 1962, 28 (6), 21-48.

Fils, D. H., & Attwell, A. A. Types of inservice meetings preferred by teachers of the educable mentally retarded. *Exceptional Children*, 1966, 33 (3), 180-181.

Fleischer, R. The counselor's role in counselor training. *Cerebral Palsy Review*, 1964, 25 (1), 8-11.

Goldman, W. J. Identification of teaching practices peculiar to special classes of educable retarded children in selected Massachusetts schools. *American Journal of Mental Deficiency*, 1959, 63, 775-777.

Harrison, J. H. The administrator looks at in-service training: A discussion. *Mental Retardation*, 1963, 1, 16-17.

Hylbert, K. W. Bachelor of rehabilitation. *Journal of Rehabilitation*, 1963, 29 (2), 23-24.

Increasing the supply of qualified rehabilitation counseling personnel in state vocational rehabilitation agencies—guidelines for action. Department of Health, Education, and Welfare, Vocational Rehabilitation Administration, Washington, D. C., January 3, 1966.

Jaslow, R. I. The division of mental retardation. *Rehabilitation Record*, 1967, September-October, 4-5.

Kent State University Follow-up Study of Participants in the Placement Training Institutes. Rehabilitation counselor perceptions of the effectiveness of placement training. Regional Rehabilitation Research Institute, University of Wisconsin, 1965, July.

Kokaska, C. J. Training volunteers for a camp program. *Mental Retardation*, 1966, 4 (2), 22-24.

Lerner, J. S. Recommended techniques for in-service education of teachers of mentally retarded children in state institutions. *American Journal of Mental Deficiency*, 1957, 62, 476-480.

Mackie, R. P., & Williams, H. M. Teachers of exceptional children. *Review of Educational Research*, 1959, 29, 395-407.

Meyers, J. S. Education and rehabilitation. *Journal of Education*, 1962, 145 (2), 1-34.

Milazzo, T. C., & Blessing, K. R. The training of directors of special education programs. *Exceptional Children*, 1964, 31 (3), 129-141.

Olshansky, S. An evaluation of rehabilitation counselor training. *Vocational Guidance Quarterly*, 1957, 5, 164-167.

Olson, J. L., & Hahn, H. R. One approach to preparing teachers of the mentally retarded. *High School Journal*, 1964, 48, 191-197.

Progress report of the rehabilitation research and training center in rehabilitation. Prepared by R. F. Heber, University of Wisconsin, 1967, October.

A proposal for a research and training center in mental retardation. Submitted by the University of Wisconsin to the Vocational Rehabilitation Administration, United States Department of Health, Education, and Welfare, 1964, March.

Stevens, G. D. A workshop approach to the training of teachers of the mentally retarded in a residential institution for the mentally deficient. *American Journal of Mental Deficiency*, 1955, 60, 278-284.

Stevens, H. A. The administrator looks at in-service training. *Mental Retardation*, 1963, 1, 13-15.

Studies in rehabilitation counselor training—communication, dissemination, and utilization of rehabilitation research information. Joint Liaison Committee of the Council of State Administrators of Vocational Rehabilitation and the Rehabilitation Counselor Educators, 1967, 5.

Warren, S. L. Internship program for rehabilitation counselors. *Journal of Rehabilitation*, 1957, 23 (3), 4-5, 20-22.

Whitten, E. B. Rehabilitation counselor education. *Journal of Rehabilitation*, 1957, 23 (3), 2, 14.

Wolinsky, G. F. Issues in curriculum development for the educable mentally retarded. *Training School Bulletin*, 1967, 64 (1), 5-11.

Wolinsky, G. F. Some considerations of the role of the teacher in a special class. *American Journal of Mental Deficiency*, 1958, 63, 415-418.

Wolinsky, G. F. Theoretical and practical aspects of a teacher education program for teachers of 'the trainable child.' *American Journal of Mental Deficiency*, 1959, 63, 948-953.

Younie, W. J. Increasing cooperation between school programs for the retarded and vocational rehabilitation services: An experimental teaching approach. *Mental Retardation*, 1966, 4 (3), 9-14.

Table 3
Etiology and General Characteristics of the Mentally Retarded

Content Areas
1. Basic Causes of Mental Retardation
a. Genetic factors
b. Environmental factors
c. Defects related to prolonged pregnancies
d. Effect of delayed inter-uterine development of a child's physical and mental ability
2. Definition and Classification of Mental Retardation
a. Intelligence and environment as related to retardation
b. Intelligence and genetic factors as related to retardation
3. Causes of Brain Injury
a. Causes of other neurological impairments
b. Learning disabilities of brain injured
4. Developmental Aspects of Mental Retardation
a. Language development
b. Auditory disorders
c. Physical development of the retarded
d. Emotional and social disturbances in the retarded
e. Behavioral disorders and mental retardation
f. Personality characteristics of the retarded
g. Psychological problems associated with mental retardation
h. Biochemical factors as related to retardation
5. Historical Survey of Mental Retardation
6. Frustration and Tolerance Levels of the Retarded
7. Role of the Medical Profession as Related to the Causes and Prevention of Retardation
8. Nutritional Factors as Related to Retardation

Table 4
Evaluation and Assessment of the Mentally Retarded

Content Areas
1. Tests
a. Types of tests used with the retarded
b. Specific tests used with the retarded
c. Effectiveness of testing procedures used with the retarded
d. Methods of evaluating and assessing the retarded
e. Standardization procedures in developing instruments for use with the retarded
2. Testing Procedures: When to Evaluate and Re-Evaluate the Retarded
3. Evaluation and Diagnosis
a. Evaluating and diagnosing the retarded
1. potential and interests of the retarded
2. functioning levels of the retarded
3. social behavior of the retarded
4. work performance (general) of the retarded
b. Prediction of vocational success and placement techniques through evaluation and diagnostic procedures
c. Evaluation and diagnosis of the—
1. auditorily and visually impaired retarded

(Continued on page 79)

Table 4 (Continued)

- 2. multiply handicapped retarded
- 3. brain damaged retarded
- 4. learning problems in the retarded
- 5. severely and profoundly retarded

d. Identification of potential delinquents in the retarded population
e. Identification of reading problems in the retarded

4. Interpretation of the Results of Diagnosis and Evaluation by Levels of Retardation
   a. The factors, sociological and behavioral, which affect decisions to institutionalize the retarded
   b. Causes of mislabeled retarded pupils (pseudo-diagnosis)
5. System Analysis of the Diagnostic Process
6. Single Disability Evaluation Centers

Table 5
Specialized Behavioral Techniques Used with the Mentally Retarded

Content Areas
1. Operant Conditioning Techniques and Training of Non-Professional Staff in Implementing Treatment Programs
2. Specific Behavioral Problems and Behavior Modification Treatment Programs
3. Use of Psychiatric Medicine in Behavioral Disorders with the Retarded
4. Use of Positive Reinforcement: Payment or Reward with the Retarded
5. Use of Negative Reinforcement with the Retarded
6. Principles of Learning Theory as Related to Training Methods in Self-Help and Socialization Skills with the Retarded
7. Motivational Techniques Used with the Institutionalized Retarded
8. Adaptive Behavior Problems in Programming for the Retarded
9. Mental Retardation and Environmental Delinquency
10. Management and Treatment Programs of the Aggressive Retarded, with Supplementary Disabilities, Epileptic
11. Establishment of Positive Attitudes by Practicing Professionals toward Treatment Programs in Use with the Retarded

Table 6
Counseling

Content Areas
1. Counseling the Mentally Retarded
a. Counseling roles and techniques for the retarded
1. role of the counselor
2. differences in counseling techniques according to level of retardation
3. group counseling techniques used with the mentally retarded
4. development of communication techniques: verbal and nonverbal, with retarded clients
5. role of the institutional vocational rehabilitation counselor in counseling with retarded clients

(Continued on page 80)

Table 6 (Continued)

6. responsibilities of the employment counselor in serving retarded clients
b. Training counselors to increase their effectiveness in the counseling process
c. Delegation of responsibility for long-term follow-up counseling needs of retarded clients
d. Specific counseling problems with the retarded
1. the delinquent retarded
2. the culturally deprived retarded
3. the specific disability groups of retarded individuals
4. the corrective institute parolees, labeled as mentally retarded
5. speech problems and their influences on the counseling situation
e. Availability of counselors to the retarded population for counseling relationships
f. Community resources available to counselors in treatment programs
2. Counseling Parents of the Retarded
a. Techniques of group counseling with parents
b. Counseling techniques in regard to—
1. vocational rehabilitation
2. placement of the retarded
3. acceptance of the child
4. realistic expectations of the child
5. training a child to capacity
6. the significance of school as well as work
7. preventive genetic measures
c. Parental involvement and cooperation in the counseling process
d. The retarded individual's effects on other family members.

Table 7

Various Care Programs Used with the Mentally Retarded

Content Areas
1. Institutional Care Programs and the Mentally Retarded
a. Institutional programs and practices oriented toward—
1. self-care activities
2. social adjustment
3. preparation for independent living
b. In-service staff training programs and practices oriented toward—
1. nursing and attendant techniques
2. nursing and attendant motivational levels
3. instruction in self-care
c. Techniques in institutional care and management of—
1. the disturbed retarded
2. the cerebral palsied retarded
3. the severely and profoundly retarded
4. adolescent retarded
5. parental contact with the institutionalized retarded
d. Description of residential life of the mentally retarded
e. Mental illness and mental retardation in an institution: integration or separation of care and treatment programs
2. Home Care and the Mentally Retarded
a. Dimensions of home care
b. Care of the infant and pre-school retarded
c. Home instruction in self-care activities with the retarded
3. Care, Treatment and Training of the Adult Retarded in Home and Community Programs

Table 8
Recreation for the Mentally Retarded

Content Areas
1. Leisure Time Activities Before and After Placement
2. Recreation and Handicraft Programs
3. Camp Programs
4. Recreation Needs and Programs for— a. Sheltered workshops b. The multi-handicapped retarded c. The severely and profoundly retarded d. The institutionalized retarded e. Rural adolescent retarded
5. Qualifications of a Recreational Staff Member
6. Clarification of Professional Responsibility for Leisure Time Activities with the Retarded

Table 9
The Retarded in Rural Areas

Content Areas
1. The Structure of Rural Programs: Grouping Techniques
2. Implementation of Rural Rehabilitation Programs
3. Rural Occupational Placement Needs and Services
4. The Establishment of Training Programs
5. Means of Obtaining Increased Services to the Rural Retarded
6. Cultural Deprivation as Related to the Rural Retarded

Table 10
Programming and Curriculum

Content Areas
1. Curricula Programming and Guides for: As Related to the Education of—
a. Pre-school educable and trainable retarded
b. Primary educable and trainable retarded
c. Intermediate educable and trainable retarded
d. Secondary educable and trainable retarded
1. reading
2. arithmetic
3. language and vocabulary
4. sex education
5. driver education
6. social skills
7. remedial programs
8. work-study programs
2. Programming and Curriculum as Related to School-Work Programs
a. Establishment of a school-work program
b. Coordination of a school-work program
c. The relation between the classroom and the work experience: classroom preparation and evaluation
d. The relation between the school administration and the work experience programs
e. Specific examples of school-work programs
f. Teaching methods for occupational adjustment programs
g. Suitable materials for vocational programs
h. Recruitment of employers for work programs
i. Job placement of the retarded
j. Legal implications of school-work programs
3. Scope and Sequence in Curricula for the Retarded
4. Suitability and Availability of Curricula Materials in Relation to Level of Retardation
5. Teaching Techniques as Related to Level of Retardation
6. Techniques in Developing the Self-Concept of the Retarded
7. The Role of the Public School in Programming for the Mentally Retarded
8. Specific Considerations in Program Planning
a. Program planning as related to the needs of—
1. the multiply handicapped
2. the special learning disability group
3. the adolescent school drop-out
4. the culturally deprived
5. the emotionally disturbed
6. the slow learners
7. the homebound retarded
8. the geriatric retarded
9. the profoundly retarded
b. Scheduling considerations
1. apportionment of time in various academic and vocational subject areas
2. appropriate use of modular scheduling
3. placement of the retarded in special and/or regular classes
c. Programming as related to the function, structure and establishment of day care centers
d. Planning a secondary program to meet the needs of the retarded
e. Planning a summer program to meet the needs of the retarded
f. The relation between physical facilities and program planning
9. Techniques in the Evaluation of Programs

Table 11
Vocational Rehabilitation

Content Areas
1. Definition of Population Being Served by Vocational Rehabilitation Personnel
2. Work Adjustment Programs a. Procedures in the establishment of various vocational rehabilitation programs b. Pre-vocational training programs in relation to— 1. development of social skills in the retarded 2. development of work attitudes in the retarded c. Techniques of on-the-job training and supervision for the retarded d. Motivational techniques as related to employment for the retarded e. Techniques of vocational evaluation and prediction with the retarded f. Employment opportunities for the retarded in relation to program planning g. Responsibility for vocational placement of the retarded h. Work adjustment program—employer relationships 1. education of the employer 2. client information useful to the employer i. Research on job permanency
3. Current Trends, Techniques, and Materials of Rehabilitative Programs
4. Rehabilitation of the Multiply Handicapped: Use of Prostheses, Etc.
5. Rehabilitation of the Severely and Profoundly Retarded: Home or Self-Employment
6. Adult Education Programs for the Retarded
7. Long-term Planning and Expectations of the Retarded
8. The Relation between Vocational Rehabilitation Training and Work Programs in Institutions
9. The Role of Vocational Rehabilitation Services in the Provision of Living Arrangements for the Retarded a. Independent living programs b. Halfway house programs c. Guest house programs d. Foster home programs
10. Development and Organization of a comprehensive Rehabilitation Center for the Retarded
11. Planning and Development of Community Services for Vocational Rehabilitation of the Retarded
12. Quality of Services in Relation to Case Load and Administrative Duties with Retarded Clients

Table 12
Sheltered Workshops and the Mentally Retarded

Content Areas

1. Development of an Extended Employment or Terminal Workshop for the Retarded
   a. Orientation and philosophy of workshops
   b. Services available in the establishment of a workshop
   c. Workshop activities and structure
   d. Management of a workshop: staff ratio with retarded clients
   e. Financial aspects of workshop programs
   f. Comparison of state and privately operated workshops
   g. Production and inventory control techniques
   h. Subcontract bidding
      1. methods of obtaining contracts
      2. vocational rehabilitation service assistance in obtaining contracts
   i. Effective marketing of products
   j. Relation between union and workshop employment
   k. Specific workshop clients problems
      1. management of social problems
      2. the place of the non-productive worker
   l. Geographical location of the workshop
2. Placement Problems with the Retarded
   a. Determining eligibility and potential of workers for specific jobs
   b. Training and programming as related to placement
      1. development of task motivation
      2. development of task persistence
   c. Type of employment as related to individual need
   d. Role of the supervisor in placement practices: supervisor-employer reelationships
   e. Vocational evaluation: its strength and weakness in relation to placement
   f. Utility of work samples with the retarded
3. Legal Aspects of workshops Employment for the Retarded
   a. Legal responsibility of employer
   b. Wage and hour law as it affects workshops
   c. Labor laws affecting workshop programs
   d. Insurance coverage of student trainees in workshops
4. The Function of the Workshop within a State or Private Institution Serving Retarded Clients
5. Methods of Facilitating the Transition from Sheltered to Competitive Employment
6. Comparison of Sheltered Workshop Programs with Training Programs for Non-Employment
7. Considerations in Terminal Employment Decisions with the Retarded
8. Requisites for Staff and Staff Training Programs in Mental Retardation
9. Tours of Sheltered Workshops
10. Case History Studies of Successful Employment Programs with the Retarded

Table 13
Administration

Content Areas
1. Multi-Disciplinary Approach in Provision of Services to the Retarded
a. Coordination of services in communities
b. Role of specific personnel in servicing the mentally retarded
1. vocational rehabilitation administrators and counselors
2. teachers
3. sheltered workshop personnel
4. guidance counselors
5. medical personnel
6. special education personnel
7. therapists
8. nurses
9. non-professional aides and volunteers
c. Communication between professional disciplines
2. Staff Training
a. Definition of personnel qualifications for work with the retarded
b. Recruitment of skilled personnel
c. Types of in-service training programs
d. Academic courses in staff training programs
e. Methods of re-evaluating teaching techniques
f. Methods of implementing new teaching techniques
g. Staff libraries in relation to personnel training
3. Administrative Practices and Techniques
a. Cooperative programs involving several government agencies
b. Role of the state department of public instruction in the development of a school program
c. Supervision for professional and non-professional personnel
d. Fostering client-centered staff communication
e. Practices in acquiring materials and funds
f. Practices in fostering creativity and resourcefulness
4. Finances
a. Federal and state funding possibilities for the retarded
b. Grant application procedures in mental retardation
c. Preparation of budgets for special education with an emphasis on less record keeping for school and vocational rehabilitation personnel
d. Financial aid for the unemployed retarded living at home
e. Production costs for competitive and sheltered employment
f. Information concerning fellowships for advanced degrees in mental retardation

Table 14
Public Relations

Content Areas
1. Methods of Encouraging Community Knowledge and Acceptance of–
a. School-work programs
b. Sheltered workshops
c. Vocational rehabilitation training programs
d. Employment of retarded individuals
2. Methods of Eliciting Community Support of, and Interest in, Programs for the Retarded
3. Use of Existing Community Resources
4. Utility Group Work Activities in the Community

Table 15
Research

Content Areas
1. Review of Current Programmatic and Longitudinal Research in Mental Retardation
2. Specific Research Areas of Interest a. Research on the mildly and moderately retarded b. Mental retardation and cultural-familial factors: core area research c. Mental retardation and genetic factors: prevention oriented research d. Mental retardation and eugenics: pros and cons e. Nutritional factors and the retarded child f. Mental retardation and vocational rehabilitation research g. Mental retardation and social work research h. Mental retardation and alcoholism i. Mental retardation and crime, welfare programs, and school drop-out rates j. Review of current University of Wisconsin research programs
3. Integration of Research into Action
4. Medical Research as Related to the Incidence of Mental Retardation
5. Techniques of Data Gathering and Processing

Table 16
Therapy

Content Areas
1. Speech Therapy and the Mentally Retarded
2. Physical Therapy and the Mentally Retarded
3. Occupational Therapy and the Mentally Retarded
4. Psychotherapy and the Emotionally Disturbed Mentally Retarded
5. Adjunctive Therapies in Hospital Programs
6. Integration and Coordination of Therapeutic Services
7. Activities as Therapy
8. Materials for Therapists
9. Group Therapy Techniques
10. Electroshock Terapy
11. Orthopedic Problems and Therapy
12. Convulsive and Epileptic Problems and Therapy
13. Pharmacology and Therapeutics

Table 17
Sociology and Psychology

Content Areas
1. Sociological Concepts Related to Mental Retardation
2. Psychological Aspects of Mental Retardation (Needs and Weaknesses in Adult Retarded)
3. Psychologist's Role in the Rehabilitation Programming for the Mentally Retarded
4. Psychiatric Problems Associated with Mental Retardation

Table 18
The Content Interests of the Single Disciplines

Vocational Rehabilitation Administrators (Content)	Vocational Rehabilitation Counselors (Content)	Educators (Content)
1. Vocational Rehabilitation	1. Vocational Rehabilitation	1. Programming & Curriculum
2. Administration	2. Counseling	2. Vocational Rehabilitation
3. Sheltered Workshop	3. Evaluation	3. Administration
4. Counseling	4. Administration	4. Sheltered Workshop
5. Programming & Curriculum	5. Sheltered Workshop	5. Etiology
6. Public Relations	6. Etiology	6. Evaluation
7. Evaluation	7. Programming & Curriculum	7. Counseling
8. Etiology	8. Public Relations	8. Public Relations
9. Recreation	9. Behavioral Techniques	9. Rural
10. Research	10. Rural	10. Behavioral Techniques
11. Behavioral Techniques	11. Care	11. Research
12. Care	12. Research	12. Recreation
13. Therapy	13. Recreation	13. Care
14. Rural	14. Sociology & Psychology	14. Therapy
15. Sociology & Psychology		15. Sociology & Psychology

(continued)

Table 18 (continued)
The Content Interests of the Single Disciplines (continued)

Psychologists (Content)	Clergy (Content)	Medical Personnel (Content)	Therapists (Content)
1. Evaluation	1. Programming & Curriculum	1. Etiology	1. Therapy
2. Vocational Rehabilitation	2. Care	2. Vocational Rehabilitation	2. Etiology
3. Programming & Curriculum	3. Recreation	3. Therapy	3. Administration
4. Behavioral Techniques	4. Etiology	4. Administration	4. Recreation
5. Sheltered Workshop		5. Care	5. Vocational Rehabilitation
6. Counseling		6. Behavioral Techniques	6. Counseling
7. Etiology		7. Research	7. Behavioral Techniques
8. Care		8. Sociology & Psychology	8. Care
9. Public Relations			9. Sheltered Workshop
10. Administration			10. Sociology & Psychology
11. Therapy			
12. Research			

(continued)

Table 18 (continued)
The Content Interests of the Single Disciplines (concluded)

Business Personnel (Content)	Sheltered Workshop Personnel (Content)	Nurses (Content)	Social Workers (Content)
1. Vocational Rehabilitation	1. Sheltered Workshop	1. Care	1. Vocational Rehabiliattion
2. Sheltered Workshop	2. Vocational Rehabilitation	2. Programming & Curriculum	2. Counseling
3. Programming & Curriculum	3. Counseling	3. Behavioral Techniques	3. Programming & Curriculum
4. Etiology	4. Programming & Curriculum	4. Sociology & Psychology	4. Etiology
5. Administration	5. Care	5. Vocational Rehabilitation	5. Sheltered Workshop
6. Evaluation	6. Rural	6. Administration	6. Evaluation
7. Care	7. Etiology	7. Counseling	7. Recreation
8. Counseling	8. Recreation	8. Etiology	8. Public Relations
9. Therapy	9. Research	9. Therapy	9. Research
10. Public Relations	10. Public Relations	10. Research	10. Administration
	11. Administration		11. Therapy
			12. Care
			13. Behavioral Techniques

Evaluation of Training: Annotated Bibliography

Alkin, M. C. Evaluating the cost-effectiveness of instructional programs. Symposium presented at the Center for the Study of Evaluation, Report No. 25, University of California at Los Angeles, May 1969.

Describes the application of a cost-effectiveness model to the evaluation of educational programs and systems. Alkin delineates the difference between cost-benefit analysis and cost-effectiveness evaluation. A model for cost-effectiveness evaluation in education is outlined, and the possible uses of this model in diverse evaluation situations are indicated.

Alkin, M. C. Towards an evaluation model: A systems approach. Center for the Study of Evaluation, Report No. 43, University of California at Los Angeles, February 1968.

Defines, discusses, and amplifies a definition of evaluation.

Alkin, M. C., Bentzen, M. M., & Grigsby III, J. E. CSE simulated evaluation exercise: Instruction guide. Center for the Study of Evaluation, Report No. 49, University of California at Los Angeles, in press.

Describes an exercise which develops evaluators' flexibility as they respond to the problems they find in actual field conditions. Directions are given for the construction and modification of evaluation designs. There is a discussion of the specific objectives of the exercise and scoring procedures for determining how successfully each objective was met.

Alkin, M. C., Bentzen, M. M., & Grigsby III, J. E. CSE simulated evaluation exercise: Materials supplement. Center for the Study of Evaluation, Report No. 50, University of California at Los Angeles, in press.

Presents necessary informational material for the conduct of an actual evaluation.

Andrews, K. Is management training effective? Part I. *Harvard Business Review*, January-February, 1957, 85-94.

Discusses the efficacy of university courses and in-service courses for employees of Humble Oil and Refining Company and Westinghouse.

These programs were evaluated in terms of the participants' reactions, faculty reactions, and company reactions.

Andrews, K. Is management training effective? Part II. *Harvard Business Review,* March-April, 1957, 63-72.

Discusses measurement, objectives, and policy. An attempt is made to derive conclusions from objective criteria in order to adequately evaluate problems of measurement.

Angell, D. J., Shearer, J. W., & Berliner, D. C. *Study of training performance evaluation techniques.* Technical Report: NAVTRADEVCEN 1449-1. Port Washington, New York: U.S. Naval Training Device Center, 1964.

Summarizes current Navy proficiency-evaluation methods. Performance evaluation in training situations which involve simulators and other complex equipment is discussed. The three variables which are important in the development of "a system of performance evaluation" are described, and an "automatic training/evaluation system" is explained.

Baker, R. L. Curriculum evaluation. *Review of Educational Research,* 1969, *39* (3), 339-358.

Reviews curriculum research and evaluation. Author points to the need for the differentiation of curriculum into definable and researchable subject matter. There is an emerging technology for the analysis and the evaluation of curriculum development which will clarify the relationship between curriculum research and evaluation activities.

Baxter, B., Taffe, A. A., & Hughest, J. F. A training evaluation study. *Personnel Psychology,* 1953, 403-417.

Compares the effects of two conference training programs for a simple position, debit insurance agent. One group was provided conference training in centralized schools while the other group was similarly trained under local supervision. Comparisons were made with regard to production, job satisfaction, termination rate, life insurance knowledge and supervisors' ratings. An analysis of results, conclusions, and recommendations is presented.

Belman, H. S., & Remmers, H. H. Evaluating the results of training. *Journal of the American Society of Training Directors,* 1958, *12* (5), 28-32.

Lists eight areas which can be evaluated. Reviews briefly some of the basic principles that ought to be observed in evaluation. The authors concluded that procedures and techniques of evaluation must be dealt with if evaluations are to be meaningful.

Besco, R., Tiffin, J., & King, D. C. Evaluation techniques for management development programs. *Journal of the American Society of Training Directors,* 1959, *13* (10), 13-27.

Discusses considerations which relate to evaluation. The evaluation process consists of definition of goals, determination of training needs, construction of the program, and evaluation of methods.

Bloom, B. S. (Ed.) Taxonomy of educational objectives: The classification of educational goals. *Handbook I: Cognitive Domain.* New York: Longmans, Green and Co., 1956.

Classifies educational goals in the cognitive domain. The major headings of the classification are knowledge, comprehension, application, analysis,

synthesis, and evaluation. Sample test items for each category are provided.

Blumenfeld, W. S. Attitude change as a criterion in training. *Training and Development Journal*, 1966, *20* (9), 26-32.

Reviews literature concerned with attitude change as a criterion in the evaluation of training. Emphasis was placed on an assessment of the adequacy of experimental design and the degree to which safe inferences can be made from results.

Buchanan, P. Evaluating the results of supervisory training. *Personnel*, January 1957, 363-370.

Demonstrates the effectiveness of a training program by evaluating the responses of superiors and subordinates of the trainees who participated in the supervisory training workshops. The reliability and validity of the testing instrument are also discussed.

Buchanan, P. C. Testing the validity of an evaluation program. *Personnel*, 1957, *34* (3), 78-81.

Judgments of a trainee's performance after training usually involve opinions of his superiors and subordinates. Such judgments are used to measure a training program's effectiveness. This article presents the findings of a study designed to test the validity of this criterion.

Caldwell, L. K. Measuring and evaluating personnel training. *Public Personnel Review*, 1964, *25* (2), 97-102.

States the purpose and efficacy of measurement and evaluation of training and describes the distinction between them. There is a need for achievement measures which avoid common statistical errors. Measurement criteria must be selected carefully, and procedure must include the before training base level of performance and standards of performance toward which training is directed.

Catalanello, R. F., & Kirkpatrick, D. L. Evaluating training programs—the state of the art. *Training and Development Journal*, May 1968, 2-9.

Reviews research on current techniques being used by business, industry, and government in the evaluation of their training programs. Evaluations included: (1) reactions of the trainees to the programs, (2) problems related to learning in a training setting, (3) evaluation of possible change in job behavior as a result of exposure to a program, and (4) results to be gained.

Clos, M. Evaluation of mental health workshops in Kentucky. *Journal of Education Research*, 1966, *59* (6), 278-281.

Reports measured attitude changes in teachers who participated in a three-week or fourth-month workshop on mental health by means of the Minnesota Teacher Attitude Inventory taken at the beginning and at the end of the workshop and repeated nine months later. Positive attitude changes were retained over the nine months, and proportionately greater attitude change was demonstrated among younger and less experienced teachers. Participants in the longer workshop showed more attitude change.

Cook, D. L. *PERT [Program Evaluation and Review Technique]: Applications in education.* Washington, D. C.: Office of Education, United States

Department of Health, Education, and Welfare, 1965.

The basic concepts, techniques, and methods of PERT are described, and the author postulates potential uses of PERT in a variety of educational and developmental activities.

Cox, J. A. Application of a method of evaluating training. *Journal of Applied Psychology*, 1964, *48* (2), 84-87.

Describes the Ward Edwards' formulation of value of training used to process data in this experimental study. Estimates of proficiency level attained, worth of a trained man in dollars, and training costs in dollars are presented. An evaluation of the technique and results of the study are reported.

Crawford, M. P. Concepts of training. In R. M. Gagne (Ed.) *Psychological principles in system development*. New York: Holt, Rinehart and Winston, 1962, 301-342.

Presents an overview of seven major steps in a training development project. These are described as: analysis of the system, analysis of the particular job, specification of knowledges and skills, determination of training objectives, construction of the training program, development of measures of job proficiency, and evaluation of the training program.

Cronbach, L. J. Course improvement through evaluation. *Teachers College Record*, 1963, *64* (8), 672-683.

Describes two functions of evaluation—course improvement and a means of judging individuals. The purpose of the former is stressed and several approaches are presented.

Cronbach, L. J. Evaluation for course improvement. In R. W. Heath (Ed.) *New curricula*. New York: Harper and Row, 1964, 231-248.

Defines evaluation as the collection and use of information for the purpose of making decisions about course improvement and about individuals. Suggestions are given for the use of a process study for evaluation of the classroom and for course evaluation.

Daugherty, W. D. Statistics: A method of influencing attitudes. *Journal of the American Society of Training Directors*, May 1960, 36-39.

The author asserts that while improved instruction is significant, the evaluation of training and analysis of results is more important. It is recommended that those who are skeptical about evaluation be persuaded by the proper demonstration of the use of statistics. A variety of techniques are presented.

Dressel, P. L. Evaluation procedures for general education objectives. *Educational Record*, 1950, *31*, 97-122.

Discusses ways of integrating evaluation and instruction. The author discusses 20 questions in which testing and instruction are interrelated.

Dressel, P. L., & Mayhew, L. B. Evaluation as an aid to instruction. In S. J. French (Ed.) *Accent on teaching*. New York: Harper and Brothers, 1954, 297-320.

The main foci of evaluation are program effectiveness, educational program improvement, and student motivation. There is discussion of the major purposes of evaluation, its relevance in general education, applica-

tion of techniques of evaluation in teaching, and guides for the use of concepts of evaluation.

Duel, H. J. Hidden indicators of training effectiveness. *Journal of the American Society of Training Directors,* March 1959, 3-5.

Recommends indirect measures of training effectiveness such as changes in turnover, output, absenteeism, grievances, rejects, etc. as vital sources of information for a comprehensive evaluation. A description is given of of Training Needs Rating Form that was developed for part of the process of data collection.

Fredricksen, N. Proficiency tests for training evaluation. In R. G. Glaser (Ed.) *Psychological research in training and education.* Pittsburgh: University of Pittsburgh Press, 1962, 323-346.

Discusses the assessment of training, including the following types of training evaluation measures: (1) soliciting opinions, (2) administering attitude scales, (3) measuring knowledge, (4) eliciting related behavior, (5) eliciting what-would-I-do behavior, (6) eliciting lifelike behavior, and (7) observing real-life behavior.

Fryer, D. H. Training, and its evaluation. *Personnel Psychology,* 1951, 29-37.

Discusses methods of identifying behavior necessary for job accomplishments. Techniques of interviewing and observing are presented. An analysis of probable attitude-knowledge-skill patterns of the job is discussed.

Gilbert, J., Campbell, H. G., & Oliver, A. E. An evaluation of interdepartmental training with objective tests. *Journal of the American Society of Training Directors,* May 1963, 46-54.

Discusses and evaluates tests developed and used in primary interdepartmental training courses. The evaluation covers a description of the student population, item analysis, and the results of testing group meetings.

Glaser, R., & Klaus, D. J. Proficiency measurement: Assessing human performance. In R. M. Gagne (Ed.) *Psychological principles in system development.* New York: Holt, Rinehart and Winston, 1962, 419-474.

Describes the role of proficiency measurement in the context of system development. Specific treatment is given to: (1) criterion vs. norm-referenced measures, (2) uses of proficiency measurement, (3) definition of the behavior to be measured, (4) sampling and the relative importance of performance components, (5) precision and relevance in proficiency measurement, (6) eliciting behavior for measurement, and (7) some applications of proficiency measurement.

Goodacre, D. The experimental evaluation of management training. *Personnel,* May 1957, 534-538.

Presents the B. F. Goodrich Company evaluation of a management training program. Attitude scales were used to evaluate (1) attitudes toward the company, (2) attitudes toward employees, (3) job satisfaction, and (4) self-confidence in dealing with a variety of situations. Achievement tests were used to evaluate knowledge gained in the program. Ratings by immediate superiors were used to evaluate job performance.

Harmon, F. L., & Glickman, A. S. Managerial training: Reinforcement through evaluation. *Public Personnel Review,* 1965, *26* (4), 194-198.

The evidence of the effects of training on job performance is discussed,

and a method of evaluation is demonstrated on a training course. The evaluation process is employed to encourage positive reinforcement of what has been learned.

How to get better trainee performance appraisals through the "critical incident' technique (training methods 10). *The Training Workshop.* Waterford, Connecticut: Bureau of Business Practice, 1967.

The lack of sufficient time to observe trainees is a frequent problem to supervisors. Noticing and recording critical incident performance by the trainees is one solution to the problem. Examples are given to illustrate both satisfactory and unsatisfactory critical incidents and an example of the form for recording them is presented. How to handle various reactions of trainees to their ratings on job performance is discussed.

Hull, T. F. What should we evaluate? *Journal of the American Society of Training Directors,* January 1959, 38-39.

Cites criteria for evaluation. On-the-job application of the training program is the goal, and a discussion of the application of evaluation is presented along with some alternatives.

Junker, E. S. Hiring, training, and evaluation of instructors. *Journal of the American Society of Training Directors,* April 1964, 23-30.

Presents and discusses important criteria relating to the hiring and training of instructors. A technique for the evaluation of instructors with the use of a rating scale is recommended, and an interpretation of evaluation results and the development of a salary schedule are presented.

Kirkpatrick, D. L. Evaluation of training. In R. L. Craig and L. R. Bittel (Eds.) *Training and development handbook.* New York: McGraw-Hill, 1967, 87-112. (*Journal of American Society for Training and Development,* November, December 1959, January, February 1960, 4 parts).

Discusses four steps considered to be critical in the evaluation process. The steps include: (1) reaction (how well the trainees like a particular training program), (2) learning (what principles, facts, and techniques were understood and absorbed by the conferees), (3) behavior (changes in performance which can be attributed to conference activity), and (4) results (reduction of costs, reduction of turnover and absenteeism, reduction of grievances, increase in quality and quantity of production, and improved morale). Emphasis is placed on the techniques which training directors can use to evaluate their own programs.

Kirkpatrick, D. L. How to start an objective evaluation of your training program. *Journal of the American Society of Training Directors,* 1956, *10* (3), 18-22.

States how an objective evaluation of a training program begins. The question, "Has the trainee learned the desired facts and principles?" is answered by four evaluative processes: use of a paper and pencil test, before and after tests of the trainee, comparison of each trainee's pre-test and post-test scores, and analysis of item change from pre-test to post-test. Explanation of the statistical findings is given.

Korb, D. How to measure the results of supervisor training. *Personnel,* March 1956, 378-391.

Reviews information about research methods, field experiments and sur-

vey techniques for developing and measuring supervisory training. Two simple plans are presented—before training and after training.

Korb, L. D. How to measure results. In U. S. Civil Service Commission Personnel Methods Series No. 4 *Training the supervisor: A guide on how to set up and conduct a supervisory training program.* Washington, D. C.: U. S. Government Printing Office, 1956.

Describes the field study which is the most likely approach to be used for scientific evaluation of training. Its components are enumerated, and instructions for the use of the evaluator and charts on supervisory training evaluator and sample plans for their application before and after training are presented. Objective recording, the use of a control group, completeness of information, and follow-up are among the procedures stressed.

Kunze, K. R. Forced choice evaluation of a training program. *Journal of the American Society of Training Directors,* August 1958, 27-32.

Describes the use of a binary measuring technique for training evaluation. A justification of this approach along with a variety of illustration scales is presented.

Lawske, C. H., Bolda, R. A., & Brume, R. L. Studies in management training evaluation. *Journal of Applied Psychology,* 1959, *43* (5), 287-292.

Reports five studies conducted to evaluate the effects of single and repeated exposures to the skit-completion method of role playing. Evaluation criteria consisted of scaled responses to a standard human relations training case in two dimensions: sensitivity and employee-orientation. Criterion responses were obtained before and after role playing in four subject groups and after the training in the fifth group. Various role playing treatment conditions and role assignments were investigated.

Lerda, L. W., & Cross, L. W. Performance oriented training—results, measurement and follow-up. *Journal of the American Society for Training Directors,* 1962, *16* (8), 12-21.

Results are measurable when a training need or problem has been explicitly delineated and when there is proper definition of the program objectives. Areas for measuring are learning and application. Included in the discussion are principles for training progress evaluation, information about evaluating results, and a scheme for measurement of results of organized training.

Lippitt, G. L., McCune, S. D., & Church, L. D. Attitudes of training directors toward the application of research to training programs. *Leadership Resources, Inc.,* Washington, D. C., or *Training Directors Journal,* March 1964.

Assesses the attitudes of training directors on a variety of issues which are critical to a comprehensive training program. In this evaluation are opinions concerning the purposes of training, methods used in evaluation, uses of evaluation material, preferences of evaluation methods, and obstacles to training, research, and research needs.

Lundberg, C. C. Attitude change and management training. *Personnel Administration,* 1962, *25* (3), 35-43.

Maintains that the concept of attitude as currently conceived by those responsible for human relations training for managers is not compatible

with the declared general training objectives. The author shows how an "attitude" actually may be a barrier to effective management action.

Lynton, R. P., & Pareek, U. Support and evaluation. *Training for development*. Homewood, Illinois: Richard D. Irwin, Inc. & The Dorsey Press, 1967, 299-321.

Discusses how post-training contracts and services help both participants and program after a training program in the transfer of learning to the work situation. Reliable criteria and evaluation procedures are necessary in the determination of what post-training contracts are indicated. They constitute the major purpose of the training and should have solid criteria and evaluation procedures. Measurement of end-of-course learning is the most salient index for evaluation of the training process and its effectiveness.

MacKinney, H. C. Progressive levels in the evaluation on training programs. *Personnel*, 1957, *34* (3), 72-77.

Describes the many ways of evaluating the results of training, and recommends recognition that some methods are better than others. Classifies evaluation procedures and criteria in descending order of acceptability.

Mahoney, T. A. Evaluation of training. *Personnel Journal*, 1960, *38* (9), 344-345.

Evaluates a management training program developed by K. Korman of the University of Minnesota and T. H. Jerdee of the University of North Carolina under the direction of T. A. Mahoney of the University of Minnesota. Course objectives are outlined. Improvement is measured by a knowledge test, a case problems test, and an aptitude scale.

Mahoney, T. A., Jerdee, T. H., & Korman, A. An experimental evaluation of management development. *Personnel Psychology*, 1960, *13* (1), 81-98.

Describes an experimental approach to the evaluation process for measuring a large industrial organization's program. The multiple objectives of the training program are recounted. Training included case analysis, group discussion, supplemental reading, and lectures. The evaluation of the course is intended to show whether course objectives were obtained and whether the "participating manager's superior" helped in the process. The course does not meet all the objectives. The "case approach and the boss-involved approach" are not completely successful and re-open the question of the value of using an analytical approach in the training.

McGehee, W., & Thayer, P. W. Evaluation of training. *Training in business and industry*. New York: Wiley, 1961.

Discusses the purpose and procedures for training evaluation. Classification of four types of evaluation measures and their respective features are delineated. All evaluation measures should be relevant, reliable, free from bias, and practical.

Messick, S. The criterion problem in the evaluation of instruction: Assessing possible, not just intended outcomes. Symposium presented at the Center for the Study of Evaluation, Report No. 22, University of California at Los Angeles, May 1969.

Discusses cognitive styles and effective reaction as important criterion variables which should be taken into account in evaluation of instruction. Messick stressed these variables because of their relevance to opinions

about the variability of human performance and the place of values in educational research.

Miller, H. L. The evaluation of education and training. *Self and service enrichment through federal training, an annex to the report of the presidential task force on career advancement.* U. S. Civil Service Commission, Washington, D. C.: U. S. Government Printing Office, 1967, 430-443.

States that specific program objectives are essential to the evaluation process. The technical contribution of evaluation is the development of tools for the measurement of program objectives. Goals of evaluation and problems in measurement are discussed.

Miraglia, J. F. Human relations training. *Training and Development Journal,* 1966, *20* (8), 18-25.

Reviews research on four categories for determining the effects of on-the-job human relations training. Evaluations of on-the-job human relations training are difficult because of a poverty of valid and reliable measures, because organizations are reluctant to cooperate in such endeavors, and because skills in supervision are slow to change.

Moon, C. G., & Hariton, T. Evaluating an appraisal and feedback training program. *Personnel,* 1958, *35,* 36-21.

States that many companies believe that evaluating the effectiveness of a training program demands elaborate, expensive procedures. This article reviews the results of a study which demonstrated the way meaningful results can be obtained from relatively simple measuring methods.

Moore, W. R. Training evaluation—it used to be so simple. *Journal of the American Society of Training Directors,* April 1964, 45-50.

Recommends a team approach to the evaluation of training. Inspectors, supervisors, trainees, and training officers need to evaluate in accordance to specified criteria. The importance of effective management in evaluation is stressed. The author suggests that only two end results be measured: quality and quantity of work on the job, and growth in the individual.

Morgan, B., Holmes, G. E., & Bundy, C. E. Evaluation in adult education. *Methods in adult education.* Danville, Illinois: The Interstate Printers and Publishers, 1960.

Classifies types of evaluation in terms of degrees of formality and precision. Informal evaluation includes observation and informal tests; semi-formal includes reliable and valid tests and surveys; formal includes re-research which involves knowledge of statistical procedures and special training. Principles of evaluation are fully discussed as are the seven steps in evaluation. A paradigm for evaluation and a short form for the evaluation of meetings are presented.

Norman, J. H. Dollars and cents evaluation of a training program. *Journal of the American Society of Training Directors,* October 1959, 32-36.

Presentation of a formula that yields an acceptable monetary appraisal of evaluation efforts.

Odiorne, G. S. A systems approach to training. *Journal of the American Society of Training Directors,* October 1965, 11-19.

Presents an argument for a systems approach to training which concerns

inputs, processes, and outputs. An explanation is given of eight kinds of training systems along with a guide for their adoption in a variety of training situations.

Pace, C. R. Evaluation perspectives. Center for the Study of Evaluation, Report No. 8, University of California at Los Angeles, December 1968.

Discusses evaluation development and the role of the evaluator in that development. The author examines the new emphasis in evaluation and the relevance of different evaluation models to other units which, in turn, were subjects for evaluation.

Parnicky, J. J. *The evaluation of institutes: A guide for measuring their impact on social work participants.* New York: National Association of Social Workers, 1966.

A guide for the aid of social work institute planning committees prepared as an auxiliary to the NASW Manual for Educational Directors and Seminar Leaders. Primary emphasis is on written techniques, with some attention given to structures and open-ended oral and observational techniques.

Pearson, J. B. (Ed.) *The analysis of short-term seminars in psychiatry for non-psychiatric physicians: A progress report for the years 1963-1966.* Boulder, Colorado: Western Interstate Commission for Higher Education, 1966.

Questionnaires from eight seminars in 1963-1964 are compared with ten seminars in 1964-1965, then analyzed for their implications for evaluation. Two questions are fundamental in all program evaluation: What is accomplished by the program? and How?

Randall, L. K. Evaluation: A training dilemma. *Journal of the American Society of Training Directors,* 1960, *14* (5), 29-35.

Presents various points of view concerning the evaluation of formal training. These views include: (1) either impossible or unnecessary, (2) the only answer, and (3) important despite unacceptable techniques. The author suggests an alternative, the "refinement" type of evaluation, which consists of three elements: verifying basic training assumptions, establishing satisfactory criteria, and implementing a meaningful evaluation design.

Remmers, H. H. *How to evaluate training programs in business and industry.* Lafayette, Indiana: Purdue University, n.d.

Presents the purpose of evaluation. The objectives of evaluation, test construction, and the determination of a test's efficacy are themes developed in four booklets from the Purdue University Press.

Rose, H. C. Evaluation of the training program. *The development and supervision of training programs.* Chicago: American Technical Society, 1964.

Explores three aspects of the evaluation process: the plan, the process, and the product. A model of a trainee evaluation questionnaire is presented, as is an explanation of the controlled research method of evaluation along with two reports of the method.

Rose, H. C. A plan for training evaluation. *Training and Development Journal,* 1968, *22* (5), 38-51.

Reviews eight elements of training and 13 techniques of obtaining infor-

mation for evaluation with a statement of their respective advantages and disadvantages. The author suggests that the results of various evaluative actions be combined to obtain the best possible total evaluation at the most reasonable costs.

Sammons, R. F. Evaluating a workshop in training skills. *Journal of the American Society of Training Directors*, July 1965, 27-33.

A description of an evaluation procedure using questionnaires.

Scheerenberger, R. C. An introduction to PERT and its application to educational programs. *The Winnower*, 1966, 2 (4), 12-22.

Describes the important features of Program Evaluation and Review Technique (PERT). It is a structure for the work organization which has generalizability to the management of programs. Its adaptability to provisions of control over such variables as time, cost, and performance is pointed out.

Schlesinger, L. Evaluating the content of multiple-skill training programs. *Personnel Administration*, 1958, *21*, 20-27.

Describes three methods of appraising the relevance of training programs to skills and knowledge required for the performance of tasks. The first criterion used was a rating scale. The second was a comparison of the time distribution of training topics and relative training emphasis with the skills and knowledge required on the job as inferred from job description data. The third criterion was selecting a small number of critically important job activities and comparing these activities with the actual content in the courses.

Schultz, D. G., & Siegel, A. I. *Post-training performance criterion development and application: A selective review of methods for measuring individual differences in on-the-job performance*. Wayne, Pennsylvania: Applied Psychological Services, 1961.

Reviews currently employed methods for on-the-job evaluation. Two necessities are: greater cohesion and congruence in assessing the field, and a cogent evaluation of measuring techniques.

Schultz, D. G., & Siegel, A. I. The rationale and application of job suitability as a basis for the evaluation of training. *Personal Psychology*, 1962, *15* (3), 261-277.

Demonstrates a technique developed by Applied Psychological Services which employed "suitability for the job" as a basis for training evaluation by: (1) describing a method for quantitatively summarizing suitability, and (2) illustrating the application of the method to data collected on technicians in four Naval jobs.

Schutz, R. E. Methodological issues in curriculum research. *Review of Education Research*, 1969, *39* (3), 354-366.

Describes good methodology as one that can distinguish between mythology and reality, strong inference and weak inference, science and technology, and between comprehensive knowledge about education and education as a conjunction domain. A search of the literature shows that previous 'methodology' was only typology-building which, in the future, will be supplanted by a form which has been produced across disciplines in the social sciences for use by people at the journeyman level of training in curriculum specialization.

Siegel, A. I., & Schultz, D. G. Evaluating the effects of training. *Journal of the American Society of Training Directors*, September 1961, 52-54.

Advocates the legitimacy of using job performance as a criterion for measuring training effects. The empirical emphasis in evaluation is defended.

Skager, R. W., & Broadbent, L. A. Cognitive structures and educational evaluation. Center for the Study of Evaluation, Report No. 46, University of California at Los Angeles.

Examines the possibilities of cognitive measures as criteria in the evaluation of instructional programs. Represents a search of the literature for descriptions of the task and for findings of relationships between cognitive measures and other influences.

Smith, R. G., Jr. *An annotated bibliography on proficiency measurement for training quality control.* Alexandria, Virginia: The George Washington University, Human Resources Research Office, 1964.

An annotated bibliography for use as a foundation for a quality control-in-training manual. One hundred and one references are listed alphabetically under five categories.

Soik, N. An evaluation of a human relations training program. *Journal of the American Society of Training Directors*, March 1958, 34-49.

Attempts to evaluate the human relations program given at the Allen-Bradley Company. Four considerations were investigated: (1) the relationship of individual characteristics to human relations knowledge, (2) the contribution of one program to increased knowledge of the trainee, (3) the relationship of class participation to increased knowledge, and (4) the need for additional training. Training consisted of guided discussion, lecture, role playing, case studies, buzz groups, and movies. Three measuring instruments were used to evaluate the effects of training: the "Supervisory Inventory on Human Relations" (S.I.H.R.), comment sheets, and "The Human Relations Questionnaire."

Stake, R. E. The generalizability of program evaluation: The countenance of education evaluation. *Teachers College Record*, 1967, *68* (7), 523-540.

An original and provocative treatment of the problem of formal evaluation which provides the foundation for the development of a plan for evaluating educational inputs rather than educational outputs. The author contends that description and judgment are necessary for the understanding of educational programs, and that there is a need for data banks for the documentation of intelligence about antecedent conditions, transactions, and intents as well as 'goals' and 'objectives.'

Stake, R. E. Testing in the evaluation of curriculum development. *Review of Educational Research*, 1968, *38* (1), 77-84.

Discusses curriculum evaluation development as distinguished from the measurement of individual aptitudes.

Stake, R. E. Generalizability of program evaluation: The need for limits. *Educational Products Report*, 1969, *2* (5), 39-41.

Discusses program evaluation which offers principles for the writing of an evaluation plan in terms of whom it serves, what to ask, what will be measured, and what the range of findings may cover.

Stephenson, H. Evaluating human relations training. *Personnel Administration*, July-August 1966, 34-39.

Compares two programs for their applicability to training in human relations. Type 1 programs use a combination of the following techniques: lecture-discussion, role playing, seminar, movies, business and computer games, and outside study. Type 1 engages the trainees for a few hours a week either at the office or a centrally designated place. A Type 2 program normally involves a concentrated one- to two-week session away from the work environment.

Stufflebeam, D. L. Evaluation as enlightenment for decision-making. Paper presented at the Working Conference on Assessment Theory sponsored by the Commission on Assessment of Educational Outcomes, the Association for Supervision and Curriculum Development, Sarasota, January 19, 1968.

Discusses the definition of the contemporary status of art in educational evaluation. Other possibilities for educational evaluation include a general definition of evaluation, an outline of four strategies for educational change activities, and strategies for the development of evaluation design structure.

Tarnopol, L. Evaluate your training program. *Training and Development Journal*, March-April 1957, 17-23.

Presents a brief justification of training evaluation.

Thisdell, R. A. Why not measure training results? *Journal of the American Society of Training Directors*, October 1959, 9-12.

Rationale for measuring training results. Two listings are given: one for a variety of ways in which training needs can be assessed and another for the ways in which training can be evaluated. The reasons training measurement is avoided are presented along with counter arguments.

Tracey, W. R. *Evaluating training and development systems.* American Management Association, Inc., 1968.

Presents a comprehensive overview of the principles and criteria for evaluation which includes philosophy and goals, management, plant and facilities, curriculum, and instructional support. A final portion of the book is concerned with a guide to rating standards.

Training methodology—Part II: Planning and administration—an annotated bibliography. Washington, D. C.: U. S. Government Printing Office, Public Health Service Publication No. 1862, 1969, 42-65.

Ninety-six annotated references on the evaluation of training are provided and categorized according to the following areas: evaluation—tests and measurement (general), evaluation—process, evaluation—specific programs, evaluation—specific methods, and evaluation—specific devices.

Trow, M. Methodical problems in the evaluation of innovation. Symposium presented at the Center for the Study of Evaluation, Report No. 31, May 1969.

Discusses innovations in the curriculum and modes of teaching and learning. Emphasis is given to the characteristics being studied and assessed, and to the social context from which they emerge.

Tyler, R. W. The functions of measurement in improving instruction. In E. F.

Lindquist (Ed.) *Educational measurement*. Washington, D. C.: American Council on Education, 1951, 47-67.

Discusses evaluation as it assists in selecting objectives, content, learning experiences, and procedures of instruction, supervision, and administration. Conditions are described that are favorable to evaluation in helping instruction.

Underwood, W. J. Evaluation of laboratory-method training. *Journal of the American Society of Training Directors*, May 1965, 34-41.

Evaluates through experimental procedures the effectiveness of sensitivity training by the T-Group method. Control and experimental groups were compared for interpersonal, personal, and nonpersonal functioning.

Vogels, D. S. Jr. An evaluation of a management training course. *Journal of the American Society of Training Directors*, January 1958, 44-51.

Evaluates management training course conducted at Griffith Air Force Base, New York. Measures of the effectiveness of training were accomplished with the use of questionnaires.

Wakely, J. H., & Shaw, M. E. Management training—an integrated approach. *Training Directors Journal*, 1965, *19* (7), 2-13.

Describes simulation in the training program of a new plan and evaluates the use of the laboratory method. Training in the program should be realistic; the staff should cooperate interdependently. There is a brief outline of all the components of the program. Four surveys were taken and four evaluations were made on four different dimensions. The results are recounted.

Walbesser, H. H. Curriculum by means of behavorial objectives. *Journal of Research in Science Teaching*, 1963, *1*, 296-301.

Reports the design and rationale for the evaluation study of the new curriculum materials for elementary school science being developed by the Commission on Science Education of AAAS. The evaluation process consisted of behavioral specifications of the curriculum objectives, their measurement by means of a checklist of competencies, and measurement of behaviors by which scientific processes are characterized. Since the aim of writing experimental material is the shaping of certain student behaviors, measures related to that behavior were included.

Weiss, C. H. Evaluation of staff training. *Welfare in Review*, 1965, *3*, (3), 11-17.

Discusses the purposes of the program of training evaluation, how results are used, and their effect upon the evaluation method. The evaluator-trainer relationship is also considered.

Wherry, R. J. The past and future of criterion evaluation. *Personnel Psychology*, 1957, *10*, 1-16.

Discusses R. M. Bellows' list of criteria for evaluation which includes: reliability, accessibility and cost, acceptability to the sponsor. In a latter portion of the article two other criteria are discussed.

Wiley, D. E. The design and analysis of evaluation studies: Comments and suggestion. Symposium presented at the Center for the Study of Evaluation, Report No. 28, May 1969.

Seeks an explicit definition of evaluation. Evaluation assessment and appraisal are made distinct from one another. The separate elements of evaluation are made clear along with their relationship to design, analysis, and measurement of evaluation.

Wilson, C. L. On-the-job and operational criteria. In R. Glaser (Ed.) *Training research and education.* Pittsburgh: University of Pittsburgh Press, 1962, 347-377.

Trainer must have feedback about on-the-job performance before there can be improvement in training. Learning theory offers a model for the requirements for three performance measurements which vary in degree of specificity. Tangible products, specific behavior elements, gross performance, and malperformance are categorized as operational performance measures. Tests constructed to resemble the regular job offer useful measures where there is a work sample whose reliability has been obtained. Ratings are not of great utility in training studies because they are not based on actual job observations and the observers do not have sufficient data for exact evaluations.

Wnuk, J. J. Evaluation of conceptual training. *Training and Development Journal,* December 1966, 38-40.

An attempt is made by the author to examine critical features of the evaluation of training programs that relate conceptual material to participants. A possible technique for evaluating conceptual training programs is presented.